Structural and Biophysical Characterisation of the Histone-like Nucleoid Structuring Protein from Enteric Bacteria

Paul G. Leonard

Department of Biochemistry and Molecular Biology
University College London, Gower Street, London

This thesis is submitted in fulfilment of the requirements of the University of London for the degree of Doctor of Philosophy

Supervisor: Professor John E. Ladbury

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

The H-NS protein is a major component of the nucleoid in enteric bacteria involved in DNA compaction and transcription regulation. The H-NS protein comprises two functional domains, an N-terminal oligomerisation domain and a C-terminal nucleic acid binding domain, separated by a flexible linker.

In this thesis, the domain architecture of the protein is investigated. The residues that form the complete N-terminal oligomerisation domain of H-NS are identified and a critical role in the formation of high order oligomeric species is established for residues Pro72 to Lys82. The oligomeric state of H-NS_{1-74} C21S and H-NS_{1-83} C21S has been determined by analytical size exclusion chromatography and analytical ultracentrifugation showing that whilst H-NS_{1-74} C21S forms a discrete homodimer, H-NS_{1-83} C21S is able to oligomerise to form tetramers and much larger protein species in a concentration dependent manner. The high order oligomerisation of H-NS is shown to be disrupted by low ionic strength and this disruption has been utilised to produce protein crystals of the complete oligomerisation domain of H-NS. Point mutations at the N-terminal and C-terminal ends of the oligomerisation domain have been identified that disrupt (R15E or E73A) or enhance (R11E or R11A) the formation of high order oligomeric species.

The DNA binding properties of the C-terminal nucleic acid binding domains of H-NS and its paralogue StpA have been directly compared by NMR. The H-NS and StpA nucleic acid binding domains bind to an AT-rich 20 base pairs DNA duplex with dissociation constants of 57 μM and 29 μM respectively.

ACKNOWLEDGEMENTS

I would like to thank Prof. John Ladbury for his support and advice throughout my time at UCL and for his infectious enthusiasm for all things H-NS. I am also grateful to John for assembling such a fantastic group of people to work with who have been constant sources of inspiration and advice to me over the last four years. Without the help and support of everyone in the Ladbury group, past and present, this PhD would not have been possible.

In particular I owe a debt of gratitude to Dr. Roger George, Dr Matt Cliff, and Dr Mark Williams for their help and training. Without Mark the NMR experiments would not have been possible and Matt and Roger have been a continuous source of guidance throughout my time in the lab. The friendship of Annika, Claire, Sunita, Tjelvar, Radwan, Shu, Sanjay, Zamal and Samad means a lot to me and has made this PhD adventure a truly enjoyable one.

An enormous thank-you must be given to Dr Gary Parkinson at the School of Pharmacy who has helped me test my protein crystals and perhaps more than anyone has suffered with me as I searched in vain for a suitable crystal to solve the structure of H-NS. Gary has gone well beyond what I have asked of him and freely given of his time and expertise, to the detriment of his own research. I will forever be grateful to him and hope that one day I will be able to return the favour.

A special mention must be reserved for Dr. Richard Harris for his critical review of this manuscript and numerous suggestions for how it could be improved.

Most importantly I would like to thank my parents who have been with me, through every high and low, from the other end of a phone line. Despite not understanding a word, they listened patiently to every biochemical problem I've encountered and provided me with encouragement and support.

This PhD project has been a considerable challenge but an enjoyable one thanks to all of my friends and family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

1D	One-dimensional
2D	Two-dimensional
Å	Angstrom
A_{280}	Absorbance at 280 nm
ε_{280}	Molar extinction coefficient at 280 nm
AUC	Analytical ultracentrifugation
B_0	External magnetic field
bp	Base pairs
Cbp	Curved DNA binding protein
Chip-on-chip	Chromatin immunoprecipitation combined with DNA microarray technology
cAMP	cyclic AMP
CD	Circular dichroism
CRP	cyclic AMP receptor protein
C-terminal	Carboxy-terminal
DNA	Deoxyribonucleic acid
Dps	DNA binding protein from starved cells
dsDNA	Double stranded DNA
FID	Free induction decay
Fis	Factor for inversion stimulation
His_6-tag	6 histidine purification tag
H-NS	Histone-like nucleoid structuring protein
$H\text{-}NS_{FL}$	Full length H-NS construct
$H\text{-}NS_{1\text{-}73}$	The subscript numbers indicate the residues from wild type H-NS present in the construct.
HU	Heat unstable protein
HSQC	Heteronuclear single quantum coherence
IHF	Integration host factor
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	kilobase
K_d	Dissociation constant
L_0	Total ligand concentration
LB	Luria Bertani
Lrp	Leucine responsive protein
MALDI	Matrix assisted lazer desorption ionisation

MRE	Mean residue ellipticity
mRNA	Messenger RNA
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
N-terminal	Amino-terminal
np	Number of points
P_0	Total protein concnetration
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein databank
PCR	Polymerase chain reaction
pI	Isoelectric point
RF	Radiofrequency
RMSD	Root-mean-square-deviation
RNA	Ribonucleic acid
S_{av}	Weight average sedimentation coefficient
$S_{av\ 20,w}$	Weight average sedimentation coefficient of the protein dissolved in water at 20°C
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
ssDNA	Single stranded DNA
StpA	Suppressor of the *Td-* phenotype
sw	Sweep width
T_2	Transverse relaxation time constant
TEV	Tobacco Etch Virus
T_M	Melting temperature
TOCSY	Total correlation spectroscopy
UPEC	Uropathogenic *E. coli*
UV	Ultraviolet
$\bar{\nu}$	Partial specific volume
ρ	Solvent density
η	Solvent viscosity

1 INTRODUCTION

The main focus of this thesis is to understand the biophysical properties of an abundant protein found in enteric bacteria called Histone-like Nucleoid Structuring protein (H-NS). H-NS has an important role in the global regulation of gene expression, particularly in response to environmental stress, and in the control of DNA packaging within the enteric bacterial cell. However, in order to understand the role of H-NS within the bacterium it is necessary to have some understanding of the bacterial nucleoid as a whole and the other proteins that are involved in packaging and regulating the DNA. An outline of the nucleoid organisation and the main proteins that control it will be given before a more detailed review of the current understanding of the H-NS protein and its paralogue StpA.

1.1 THE NUCLEOID

In bacteria the DNA is not contained in an envelope enclosed nucleus, as found in eukaryotes, but is present in the cytoplasm in a structure called the nucleoid. For *E. coli*, the 1.6 mm long chromosomal DNA molecule must be contained within a cell that is 4 µm long and 1 µm wide requiring the DNA to be compacted. Several different states of overall nucleoid structure have been observed by either indirect imaging of green fluorescent protein tagged RNA polymerase (Cabrera and Jin, 2003) or electron micrographs of *E. coli* cells under different growth conditions (Frenkiel-Krispin et al., 2004). During exponential growth the nucleoid adopts a dynamic extended conformation consisting of 50-400 negatively supercoiled DNA loops with an average size of 10 kb (Luijsterburg *et al.*, 2006). 40 nm nucleoid fibers have been observed by atomic force microscopy which are supersolenoided into 80 nm fibers of compacted DNA (Yoshimura et al., 2004). Under rapid growth conditions the nucleoid becomes slightly compacted relative to moderately growing cells. It has been suggested that this compaction is driven by the supercoiling effect of RNA polymerase transcription at rapidly transcribed genes. In low nutrient environments however, the cells enter stationary phase and time dependent electron micrographs show a transition in the DNA chromosome to several coiled toroid structures which then assemble into crystalline arrays within the cytoplasm (Frenkiel-Krispin et al., 2004) Figure 1-1. These structures will limit protein expression by reducing the accessibility of the DNA for RNA polymerase and other proteins required for transcription.

Figure 1-1. Changes in nucleoid structure dependent on the *E. coli* cell growth rate. During rapid exponential growth important operons encoding tRNAs and rRNAs (red circles) are brought into close proximity in putative transcription factories where the RNA polymerase (green circles) is able to rapidly transcribe these genes (Cabrera and Jin, 2003). At slower growth rates the nucleoid assumes a more extended conformation. Under low nutrient conditions the cell enters stationary phase and the nucleoid adopts a more compact structure consisting of several interconnected toroids of DNA. After prolonged periods of stationary phase these toroids form regularly spaced arrays within the cytoplasm (Frenkiel-Krispin et al., 2004). This figure has been reproduced from a bacterial chromatin review article by Travers A. and Muskhelishvili G. (Travers and Muskhelishvili, 2005)

The DNA packaging needs to be dynamically controlled so that regions of the DNA can be released and unwound for DNA replication or transcription processes to occur, when required by the cell. Indeed the topologically discrete domains of the nucleoid have dynamic boundaries, with no apparent sequence specificity, that can vary within a population of cells (Deng et al., 2004; Higgins et al., 1996). The packaging of DNA in bacteria is a complex process involving several different mechanisms including cellular confinement, macromolecular crowding, DNA supercoiling and nucleoid associated protein interactions.

Cellular confinement is the most immediately obvious mechanism of DNA packaging. The limited size of the cell envelope prevents the DNA from forming an extended conformation. However this only has a small effect on the size of the nucleoid as experiments using lysozyme to disrupt the cell envelope have shown. In these experiments there is a slight enlargement of the volume of the nucleoid but the overall shape of the nucleoid is retained.

Macromolecular crowding can have either a direct or indirect effect on the compaction of the DNA chromosome. Macromolecular crowding and its ability to cause DNA compaction, was first demonstrated by the use of the neutral polymers polyethylene oxide and polyvinylpyrrolidinone to compact T7 bacteriophage DNA without directly interacting with the DNA itself (Lerman, 1971). With more than 300 mg of RNA and protein per mL in the bacterial cytoplasm there is a significant crowding effect on all of the molecules within the cell. This crowding will favour the more compact conformations of the DNA itself whilst additionally encouraging the binding of nucleoid associating proteins that further package the DNA (Zimmerman, 2006). Acting conversely to the compaction of the DNA is the effect of co-translational production of secretory proteins or inner membrane proteins. Chromosomal loops encoding these proteins become transiently and indirectly anchored to the membrane by co-translational export machinery resulting in an extended conformation of the chromosome for these sections of DNA.

DNA supercoiling is the result of a DNA gyrase catalysed process that passes a segment of DNA through a single strand break in the DNA to introduce a negative superhelical tension into the DNA molecule. This causes the DNA to form plectonemic or solenoidal tertiary structures that reduce the volume of the DNA (Woldringh *et al.*, 1995). In *E. coli* mutants with reduced DNA gyrase activity or after exposure to oxolinic acid (a DNA gyrase inhibitor) the cellular DNA has a reduced level of supercoiling and a less compact structure (Steck *et al.*, 1984). Counteracting the action of DNA gyrase

are the topoisomerase enzymes; topoisomerase I, III and IV, which reduce the level of negative supercoiling in the DNA.

There is another group of abundant proteins in the bacteria cell, that play an important role in controlling the structure of the nucleoid, called nucleoid associated proteins (NAPs). These proteins share superficial similarities to eukaryotic histone proteins in terms of their abundance, basicity, low molecular weight and DNA binding properties but are structurally distinct from eukaryotic histone proteins.

1.2 NUCLEOID ASSOCIATED PROTEINS (NAP'S)

The nucleoid associated proteins can be divided into two categories; those proteins that have DNA sequence specificity and bind to specific locations of the bacterial chromosome (they include integration host factor (IHF), curved DNA binding protein A (CbpA), curved DNA binding protein B (CbpB), leucine responsive protein (Lrp)) and those proteins that do not recognise a particular DNA sequence and are located relatively uniformly throughout the nucleoid (such as histone-like protein from *E. coli* strain U93 (HU), histone-like nucleoid structuring protein (H-NS), suppressor of the Td^- phenotype (StpA), factor for inversion stimulation (Fis) and the DNA binding protein from starved cells (Dps) (Azam et al., 2000)). As H-NS and StpA are the principal proteins of interest in this report they will be discussed in more detail later in this chapter. The intracellular concentration of the nucleoid associated proteins varies in a growth phase dependent manner and thus the ratio of these proteins relative to each other will change and with respect to the concentration of DNA in the cell. Immunofluorescent imaging has been used to determine the location of these proteins with respect to the DNA in *E. coli* cells at different cell densities (Azam et al., 2000) along with quantitative western blot analysis to determine the average number of molecules present in the cell for each protein (Table 1-1) (Azam et al., 1999). In exponential growth HU and Fis are the main nucleoid associated proteins found in the *E. coli* cell. Upon entry into stationary phase the level of Fis in the cell decreases significantly whilst the intracellular concentration of Dps increases until the Dps protein is the dominant nucleoid associated protein in the cell. Dps and HU homologues have been identified in a wide range of bacteria including the alpha gamma delta and epsilon subdivisions of proteobacteria, firmicutes and actinobacteria (Ohniwa et al., 2006) whereas Fis, H-NS and StpA homologues have only been identified in enteric bacteria and a few other closely related species (Dorman et al., 1999; Tendeng and Bertin, 2003; Ohniwa et al., 2006). In this study the concentration of StpA present during

Table 1-1 The growth phase dependent changes in protein expression of nucleoid associated proteins determined by quantitative western blotting from whole cell lysates (Azam et al., 1999). *Previous studies suggest StpA is only present in the cell upon entry into stationary phase whereas H-NS is maintained at a relatively constant level with respect to the cellular DNA concentration (Free and Dorman, 1995).

exponential growth was higher than the concentration of H-NS present. However, experiments by other groups to determine the cellular concentration of H-NS and StpA showed that H-NS was present at a relatively constant ratio with respect to the amount of DNA present in the cytoplasm but StpA was only present at low levels during exponential growth in rich media (Free and Dorman, 1995; Sonnenfield et al., 2001). Under stress conditions the level of StpA was found to increase (Free and Dorman, 1995). The disparity between these findings could be caused by cross reactivity of the StpA antibody for H-NS so that the level of StpA present in the cell was over estimated.

1.2.1 Histone-like Protein from E.coli Strain U93 (HU)

HU is the main nucleoid associated protein responsible for DNA packaging during exponential growth of bacteria. Most bacteria have a homodimeric HU that consists of either two HUα or two HUβ subunits. Enteric bacteria have both α and β homologues (70% homology) which can additionally interact to form a hetero-dimeric HUαβ molecule (Pinson et al., 1999). The level of HUα and HUβ in the *E. coli* cell is growth phase dependent with a reduction in the amount of protein present, from 50000 copies per cell to 15000 copies per cell, upon entry into stationary phase (Table 1-1). Overexpression of HU *in vivo* does not result in a discernable increase in nucleoid compaction, unlike the effect of H-NS overexpression (McGovern *et al.*, 1994). It is thought that HU and H-NS act as antagonists for each other with HU counteracting the compacting effect of H-NS binding to the DNA with the higher ratio of HU to H-NS during log phase growth helping to reduce H-NS mediated compaction and allowing increased levels of mRNA transcription than those required during stationary phase when the HU and H-NS intracellular levels are more equivalent. This antagonism has been demonstrated *in vivo* using atomic force microscopy imaging of relaxed circular plasmid DNA in the presence of HU or H-NS (Dame and Goosen, 2002). The structure of HU (Swinger et al., 2003) and its close DNA sequence specific homologue IHF (Lynch et al., 2003) have been solved by X-ray crystallography revealing a predominantly α-helical core, responsible for dimerisation, with two β-ribbon arms extending out from it. The β-ribbons wrap around the minor groove of DNA and proline 63 from each protomer chain intercalate between the bases of the DNA causing the DNA to bend (Figure 1-2). Co-crystal structures of HU bound to DNA display differing bending angles from 105° to 140° depending on the DNA substrate and crystallisation conditions used. Atomic force microscopy imaging of HU-DNA complexes showed a broad distribution of angles from 0° to 180° (van Noort J. *et al.*, 2004) when a relatively

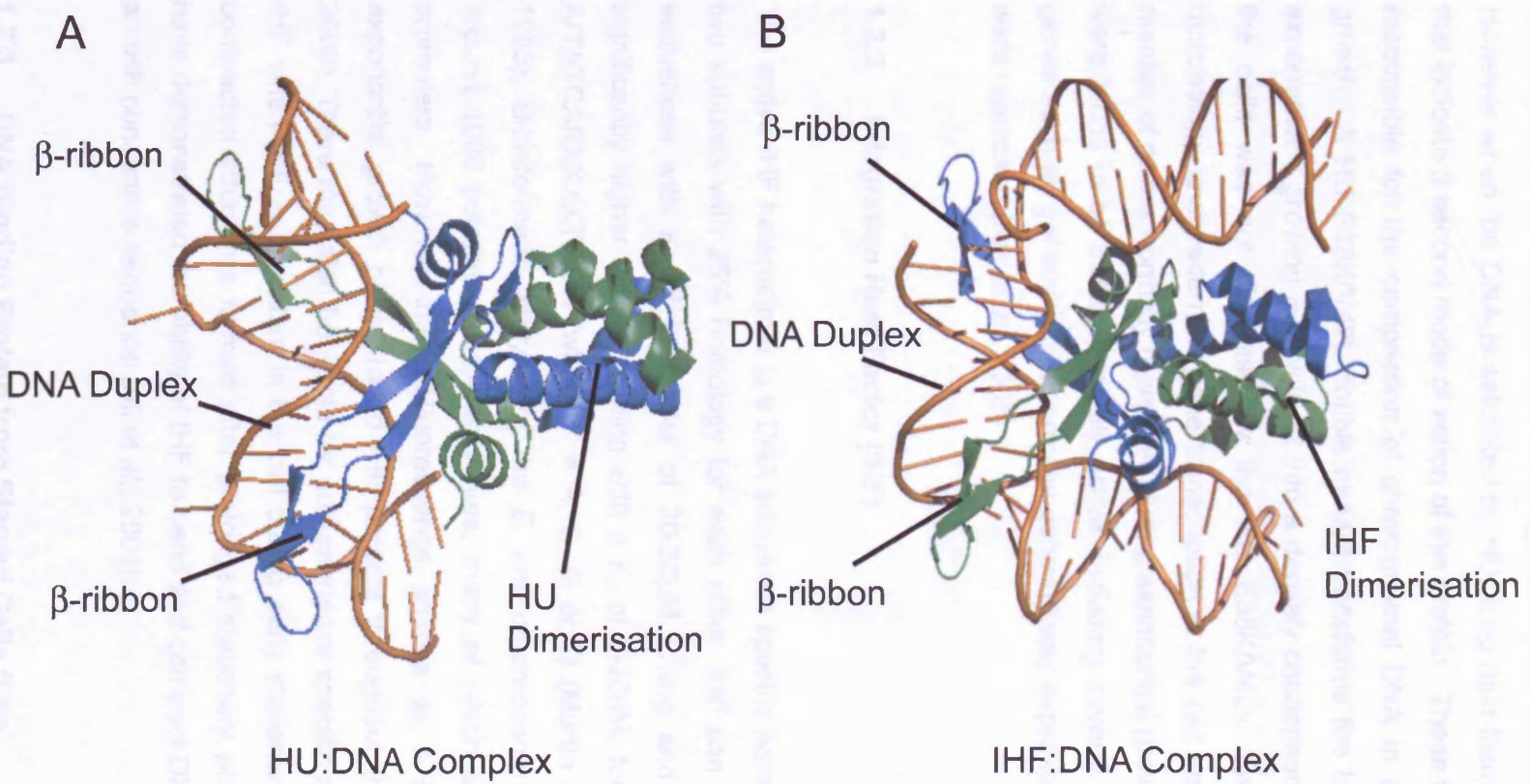


Figure 1-2. The X-ray crystal structure of HU bound to a DNA duplex (A) (Swinger et al., 2003) and IHF bound to a DNA duplex (B) (Lynch et al., 2003). Protein databank codes 1P78 and 1OWG respectively. One protomer is shown in turquoise and the other protomer is shown in green. The DNA duplex is coloured orange. The two β-ribbons from the HU or IHF dimer wrap around the minor groove of the DNA duplex. These figures were generated using Pymol (Delano, 2006).

low molar ratio of protein to DNA base pairs was used (1 HU dimer per 92 DNA base pairs). Thus when individual dimers of HU bind to DNA they can cause bending of the DNA at the point of HU binding. It is thought that interactions with the surface of the core region of HU have a role in dictating the degree of bending of the DNA molecule. However when the DNA is saturated by HU, long rigid filaments are observed *in vitro* that indicate a second mode of action of this protein. These filaments are thought to be responsible for the compaction of chromosomal DNA in *E. coli* during exponential growth. A HU E38K/V42L double mutant transforms the loosely packed nucleoid of exponentially growing *E. coli* K-12 into a densely condensed structure. The growth of the cells was not disrupted by the HU E38K/V42L double mutant but the cell morphology and pattern of gene transcription in the cell was significantly altered. A number of genes normally repressed during exponential growth in laboratory conditions were found to be transcriptionally active, including several virulence genes, whereas genes such as *gal* which are normally constitutively expressed under these conditions were repressed (Kar et al., 2005).

1.2.2 Integration Host Factor (IHF)

The active IHF heterodimer is a DNA sequence specific homologue of HU formed from two subunits with 25% homology for each other. IHF can bind to non-specific DNA sequences with K_d of the order of 20-30μM (Yang and Nash, 1995) but it has significantly higher affinity, binding with a K_d of 2-20nM, for its consensus sequence A/TATCARXXXXTTA/G (where X = A, C, G or T) (Murtin et al., 1998; Wang et al., 1995). Bioinformatic analysis of the *E. coli* chromosome sequence has identified around 1000 putative IHF binding sites; many of which are located close to gene promoters. However immunofluorescence studies at a single time point during exponential growth phase found IHF present throughout the nucleoid (Azam et al., 2000). There may also be a role for non-sequence specific compaction of the DNA by IHF when IHF is abundant in the cell during early stationary phase, prior to the Dps compacted structures formed after prolonged stationary phase. *In vitro* experiments have demonstrated the ability of IHF to bend and compact DNA even in the absence of an IHF consensus sequence (Ali *et al.*, 2001).

1.2.3 DNA Binding Protein from Starved Cells (Dps)

Dps is the most abundant nucleoid-associated protein in stationary phase *E. coli*. In a low nutrient environment a bacterium needs to be able to shut down non-essential cell

processes to reduce its energy requirement. Dps is responsible for the reorganisation of the chromosomal DNA into a more compact structure of toroids (after 24 hours of stationary phase) which then build up to form a crystalline lattice after prolonged stationary phase (48 hours) (Figure 1-1) (Frenkiel-Krispin et al., 2004). This more compact DNA structure limits the levels of transcription and protein production within the cell. Several X-ray crystallography structures of Dps have been solved for hexameric or dodecameric complexes (Grant et al., 1998). Each Dps subunit consists of a core four helix bundle made up of helices 1, 2, 4 and 5 with the short third helix exposed on the outside of the protein (Figure 1-3). However, the mode of DNA binding is unclear. The surface of the dodecamer sphere is predominantly negatively charged and would preclude any interaction with DNA. It has been suggested that pores formed when the dodecamers come together in a hexagonal 2D array contain a positively charged patch, formed from the N-terminal lysine residues of three neighbouring dodecamers in the array, which is responsible for DNA binding (Frenkiel-Krispin et al., 2004). Atomic force microscopy studies in the presence of Dps reveal a tightly packed nucleoprotein stucture where the E. coli chromosome is more tightly packaged during stationary phase relative to the sturcutre of the nucleoid observed during exponential growth (Yoshimura et al., 2004).

1.2.4 Factor for Inversion Stimulation (Fis)

Fis is a 22kDa protein (Johnson et al., 1986) consisting of four α-helices. Helix 1 and 2 from two Fis protomers interact to form the hydrophobic core of a homodimer that is the active form of the protein. (Figure 1-4) Helices 3 and 4 adopt a helix-turn-helix motif that is involved in DNA binding via the major groove of the DNA. The relative position of the two helix-turn-helix motifs in the homodimer dictate that the DNA will be bent upon binding of a Fis homodimer. DNA bends of 50° to 90° have been observed after incubation with Fis (Pan et al., 1996) and multiple Fis binding events can act in concert to create DNA bends up to 160° (Travers and Muskhelishvili, 2005). Fis recognises the DNA consensus sequence (G/T)NN(C/T)(A/G)NN(A/T)NN(C/T)(A/G)NN(C/A) (where N is any nucleotide) although there may be a role for non-DNA sequence specific Fis binding events in early exponential phase when the concentration of Fis within the cell is highest. Fis can act as an activator (Falconi et al., 2001) or repressor (Schneider et al., 1999; Xu and Johnson, 1995) of transcription. The position and number of Fis binding sites relative to a promoter may be important in determining the local conformation of the DNA around a promoter region and thus the accessibility of this DNA for RNA polymerase or transcription repressors, such as H-NS. Fis has been

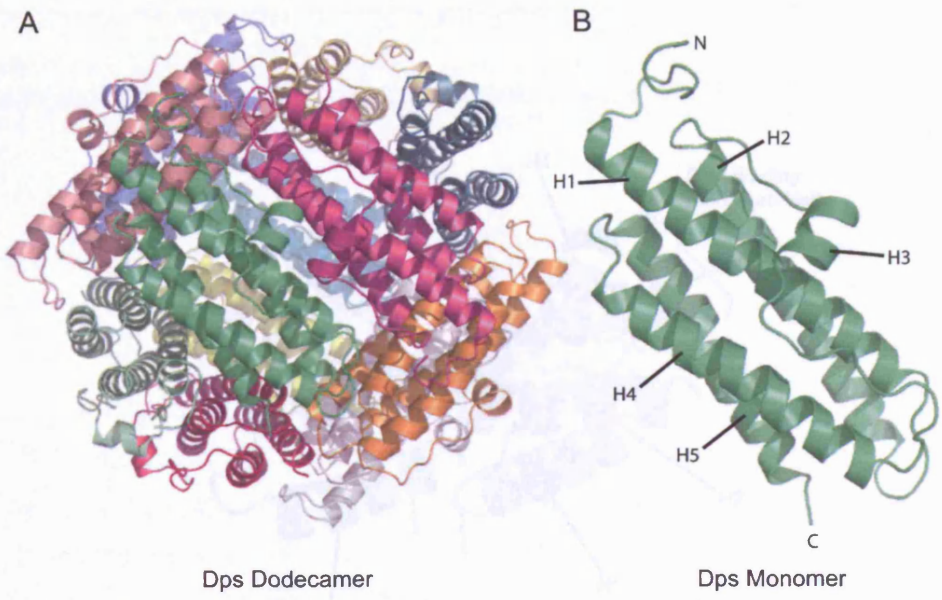


Figure 1-3. A. The X-ray crystallographically determined structure of a Dps dodecamer (Grant et al., 1998). Each of the twelve Dps monomeric subunits has been coloured with a different colour. B. One Dps molecule showing the four helix bundle formed from helices H1, H2, H4 and H5 and the relative position of the short third helix H3 that is required for DNA binding. These figures were generated using Pymol (Delano, 2006).

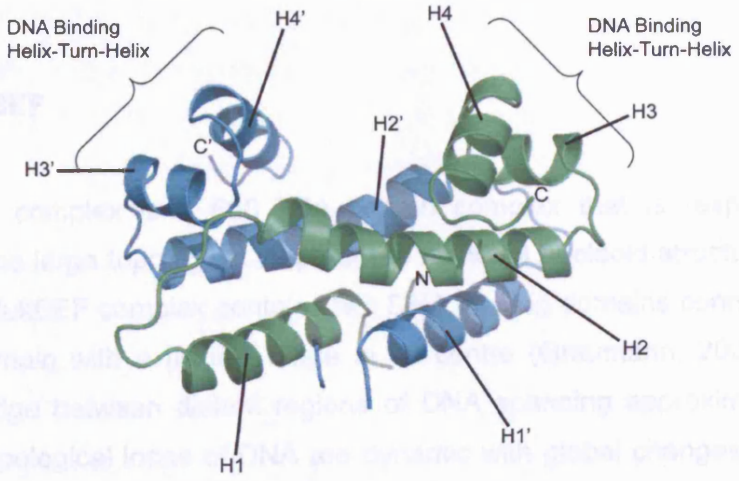


Figure 1-4. The X-ray crystallography structure of a Fis homodimer showing the two protomer chains coloured green and cyan respectively (PDB Code 1FIA) (Kostrewa et al., 1992). The core of the Fis homodimer is formed by α-helices 1 and 2 from each protomer chain. α-helices 3 and 4 form a helix-turn-helix DNA binding motif. This figure was produced using Pymol (Delano, 2006).

implicated as a global sensor of DNA topology and affects the superhelicity of the DNA by both direct interaction with the DNA and it's role as a repressor of the DNA gyrase genes *gyrA* and *gyrB* (Schneider et al., 1999). An increase in the level of negative supercoiling of the DNA results in elevated levels of Fis protein in the cell. Fis and H-NS binding sites within the *E. coli* chromosome are often found within close proximity to each other (Grainger *et al.*, 2006) and can acts as antagonists with respect to each other (Falconi et al., 2001) or in concert to enhance transcription repression (Browning et al., 2000). Changes in the ratio of the nucleoid-associated proteins with respect to both each other and to DNA will affect both the global structure of the DNA and the local structure of the DNA to influence gene expression in the bacterial cell.

1.2.5 MukBEF

The MukBEF complex is a 600 kDa protein complex that is responsible for the formation of the large topological loops of the bacterial nucleoid structure (Case et al., 2004). The MukBEF complex contains two DNA binding domains connected by a long coiled-coil domain with a flexible hinge at its centre (Graumann, 2001) allowing the protein to bridge between distant regions of DNA spanning approximately 170 base pairs. The topological loops of DNA are dynamic with global changes in the nucleoid structure regulating the expression of genes in the bacterium.

1.3 HISTONE-LIKE NUCLEOID STRUCTURING PROTEIN (H-NS)

1.3.1 Overview of H-NS

The Histone-like nucleoid structuring protein (H-NS) was first identified in E, coli nucleoid fractions (Jacquet et al., 1971) and first isolated from preparations of *E. coli* chromosomal DNA in 1977 (Varshavsky *et al.*, 1977). H-NS is a major component of the bacterial nucleoid, with tens of thousands of molecules per cell (Azam et al., 1999; Free and Dorman, 1995). It has been suggested that H-NS remains at a relatively constant ratio with respect to DNA within the cell (Free and Dorman, 1995) except in conditions of cold shock when the H-NS level increases three to four-fold (Atlung and Ingmer, 1997). Overexpression of H-NS leads to arrested cell growth of *E. coli* (McGovern *et al.*, 1994), immediate inhibition of transcription and translation, and to a lesser degree reduced DNA replication and cell wall synthesis. Immunoelectronmicroscopy studies of cells where wild type H-NS was overexpressed

revealed a highly compact, spherical nucleoid (Spurio *et al.*, 1992). Null mutants of H-NS in *S. typhimurium*, where only one H-NS family protein is present are non-viable unless mutants of either σ^S (a deletion mutant to remove residues Tyr61 to Ile65 in the σ^S protein), the stationary phase RNA polymerase sigma factor, or PhoP, the virulence gene regulator, are also introduced (Navarre et al., 2006). In *E. coli*, *hns* mutants and *hns stpA* double mutants are viable although they grow more slowly than the wild type cells (Zhang et al., 1996). Consistent with its role as a nucleoid structuring protein H-NS is able to alter DNA topology and constrain negative supercoils (Owen-Hughes et al., 1992; Tupper et al., 1994).

Combined with its function in structuring the nucleoid, H-NS acts as a pleiotropic regulator of transcription. Numerous phenotypes have been associated with mutations of the *hns* gene, including de-repression of β-glucoside metabolism (Defez and De, 1982), increased resistance to low pH and to high osmolarity (Hommais et al., 2001) and a loss of motility (Soutourina *et al.*, 2002). Gene expression profiling using DNA microarrays identified about 250 genes that are under the direct or indirect regulation of H-NS. 80% of these genes showed increased levels of mRNA transcription in an H-NS null mutant strain compared to the wild type strain (Hommais et al., 2001) supporting the current view that H-NS is a global repressor of transcription (Dorman, 2004).

The *hns* gene has been mapped to the 27.5min region of the *E. coli* chromosome (Goransson *et al.*, 1990) and encodes a 15.6 kiloDalton (kDa) polypeptide of 137 amino acids. H-NS consists of two well conserved domains: an N-terminal oligomerisation domain and a C-terminal nucleic acid binding domain with a short flexible linker between the two domains (Smyth et al., 2000). Wild type H-NS protein isolated from *E. coli* cells and analysed by mass spectrometry showed that the N-terminal methionine was cleaved in a post-translational modification step (Shindo et al., 1995). However, there is inconsistency within the literature regarding the numbering of the amino acids in H-NS. Some groups count the N-terminal methionine (present prior to post translational cleavage) as the first amino acid whereas other groups start with serine as the first amino acid. For the purpose of this report all amino acid numbering for H-NS starts with methionine as the first amino acid.

1.4 PHYLOGENETIC DISTRIBUTION OF H-NS

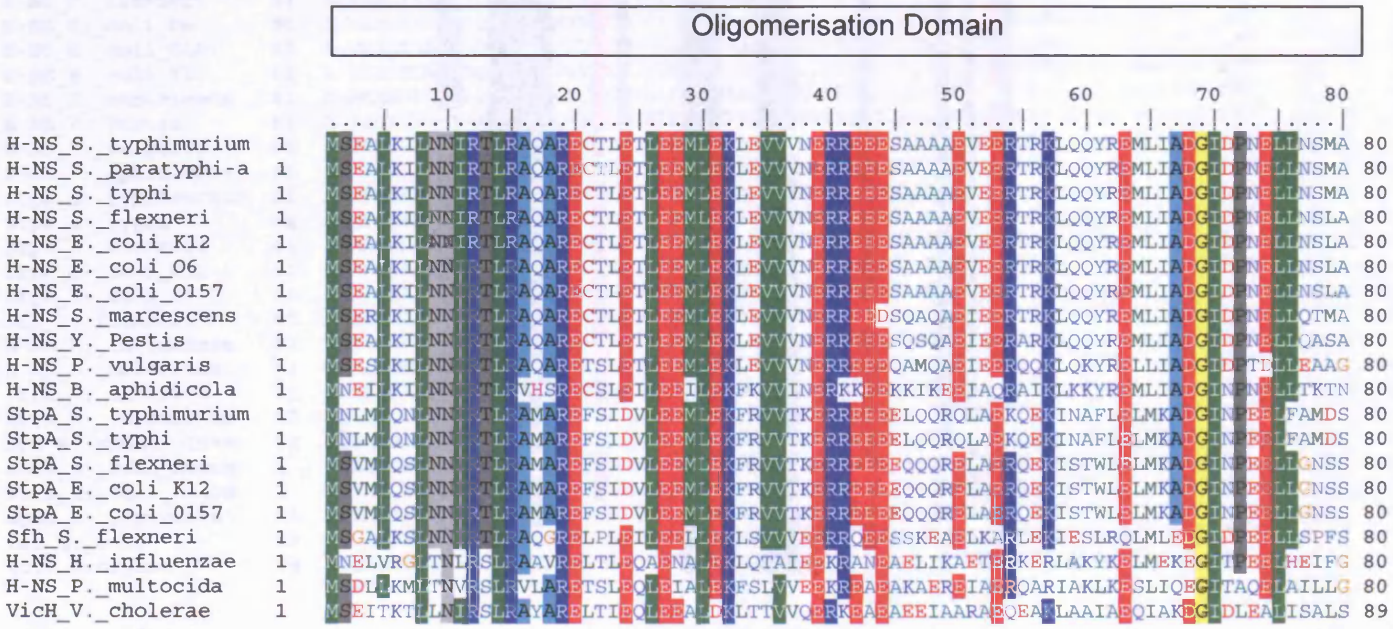
Through sequence alignments or *in vivo* complementation studies, H-NS homologues

have been identified in more than seventy Gram-negative bacteria. (Tendeng and Bertin, 2003). However H-NS homologues appear to be limited to the α, β and γ subdivisions of proteobacteria, as defined by the 16S rRNA classification system. No H-NS homologue has yet been identified in Gram-positive bacteria (Tendeng and Bertin, 2003). By contrast, homologues of another nucleoid associated protein, HU, have been identified in numerous prokaryotes as well as eukaryote organelles (Oberto *et al.*, 1994). A sequence alignment of H-NS homologues is shown in Figure 1-5.

Sequence comparisons of all known H-NS-like proteins show a high degree of conservation within the C-terminal nucleic acid binding domain but much greater variation in the N-terminal oligomerisation domain. The VicH protein from *Vibrio cholerae* (Bertin et al., 1999), and both StpA and MdbA proteins from *E. coli* (Dorman et al., 1999) show significant homology to H-NS in both the N-terminal and C-terminal regions. StpA in particular shares 57% sequence identity with H-NS and shares many of the functional characteristics of H-NS including the preference for binding to curved DNA sequences, the ability to constrain negative supercoils (Cusick and Belfort, 1998). MdbA from uropathogenic *E. coli* (UPEC) was originally identified to have homology just to the N-terminal domain of H-NS (Cusick and Belfort, 1998) but a single base pair insertion at nucleotide 405 (that could be a correction for a sequencing error) produces an open reading frame with homology to full length H-NS (Dorman et al., 1999).

A truncated form of H-NS, called H-NST, encoding just the N-terminal oligomerisation domain has been identified in enteropathogenic *E. coli* (EPEC) that displays a dominant negative effect on the repression of the *proU* operon and *bgl* gene expression (Williamson and Free, 2005). The *proU* and *bgl* promoters have been shown to be repressed by H-NS (Free et al., 2001; Williams et al., 1996; Ueguchi and Mizuno, 1993). The dominant negative effect is thought to be due to the formation of a heteromeric complex with wild type H-NS that has a lower affinity for DNA compared with H-NS oligomers. Therefore H-NST disrupts H-NS function as a transcription repressor. Another truncated H-NS-like protein has been identified in uropathogenic *E. coli* (UPEC) but ths protein has only a weak dominant negative effect on H-NS mediated *proU* repression. Co-purification studies showed that this UPEC H-NST protein had a lower affinity for wild type H-NS than the EPEC H-NST. A V16A point mutation to the UPEC H-NST protein was able to confer the EPEC H-NST phenotype to this protein.

Several proteins that contain the nucleic acid binding domain of H-NS fused to unrelated amino-terminal regions have also been identified including BpH3 from *B.*


Oligomerisation Domain
10
20
30
40
50
60
70
80
H-NS_S._typhimurium 1 80
H-NS_S._paratyphi-a 1 80
H-NS_S._typhi 1 80
H-NS_S._flexneri 1 80
H-NS_E._coli_K12 1 80
H-NS_E._coli_O6 1 80
H-NS_E._coli_O157 1 80
H-NS_S._marcescens 1 80
H-NS_Y._Pestis 1 80
H-NS_P._vulgaris 1 80
H-NS_B._aphidicola 1 80
StpA_S._typhimurium 1 80
StpA_S._typhi 1 80
StpA_S._flexneri 1 80
StpA_E._coli_K12 1 80
StpA_E._coli_0157 1 80
Sfh_S._flexneri 1 80
H-NS_H._influenzae 1 80
H-NS_P._multocida 1 80
VicH_V._cholerae 1 89

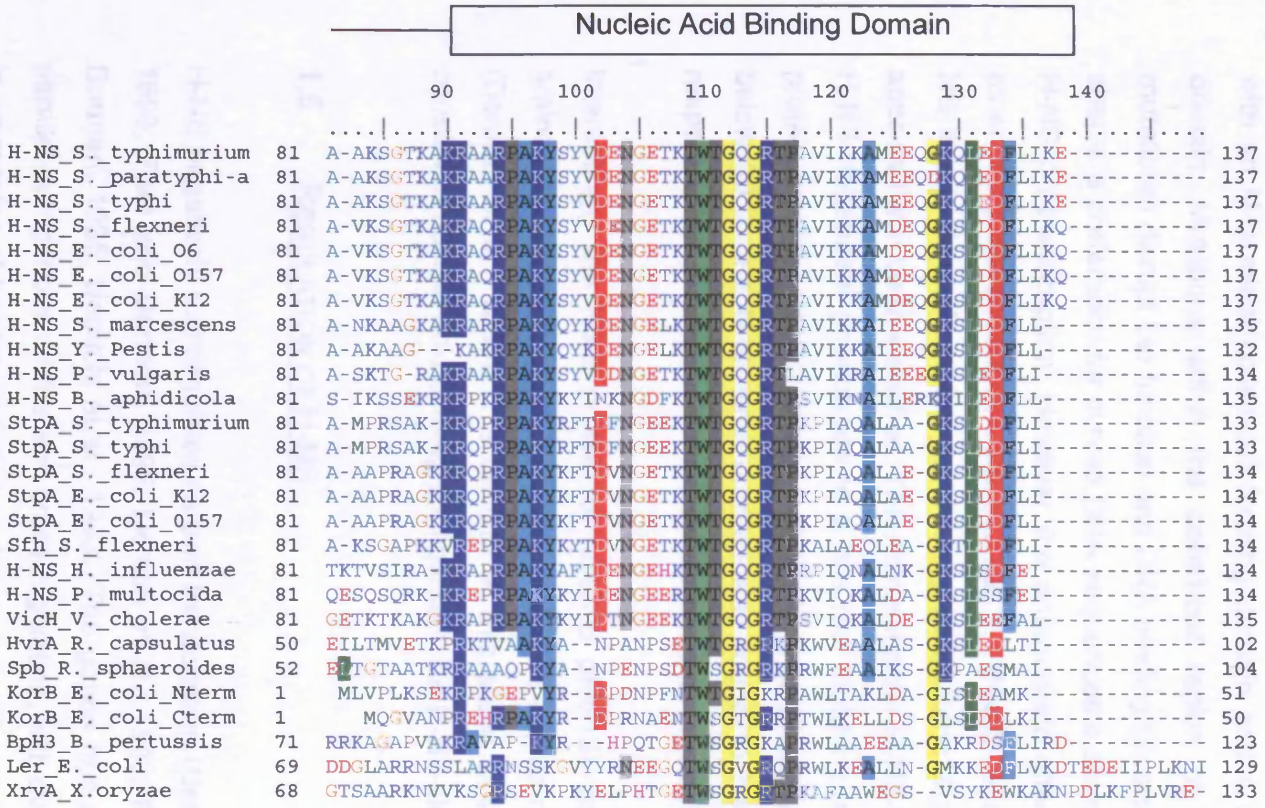


Figure 1-5. Primary Sequence Alignment of selected members of the H-NS like family of proteins. The amino acid residues are numbered with respect to *S. typhimurium* H-NS. Identical residues that are present in 75% or more of the aligned sequences have been shaded in vertical blocks. The alignment was generated using Blast2.0 sequence alignment search engine and edited using Bioedit 7.0.5.2 (Hall, 1999). The default colour scheme from Bioedit is shown with the shaded amino acids signifying >75% sequence identity in the sequences shown. The oligomerisation and nucleic acid binding domains are indicated. The * symbols below the sequence alignment indicate the putative heptad repeat residues that form the coiled-coil interactions.

pertussis, Spd from *R. sphaeroides*, XrvA in *Xanthomonas oryzae* and HvrA from *R. capsulatus* (Bertin et al., 1999). Despite the much lower amino acid sequence conservation within the N-terminal domains of these proteins, with respect to the C-terminal nucleic acid binding domains, there is functional conservation as these domains are predicted to form coiled-coil interactions (Dorman et al., 1999; Tendeng et al., 2003). Additionally the LEE encoded regulator protein (Ler) is a transcriptional activator of the H-NS repressed LEE pathogenicity island of EPEC. H-NS represses the expression of the LEE pathogenicity island proteins, including Ler, at 27°C but this transcription repression function is disrupted by a change in temperature to 37°C (Umanski *et al.*, 2002). The Ler protein displays common domain organisation to H-NS, with an N-terminal coiled-coil forming domain and a C-terminal nucleic acid binding domain. Mutations within the coiled-coil region of this protein (I26R and L29R mutations) disrupt Ler function and DNA binding (Sperandio *et al.*, 2000). Although Ler shares a preference for curved DNA sequences it antagonises the repressive effect of H-NS on transcription. However the antagonistic effect of Ler was not observed at other H-NS-dependent operons such as *bgl* (Haack *et al.*, 2003). More recently two H-NS like proteins from the psychrotrophic bacteria *Acinetobacter* and *Psychrobacter* species have been identified. The Acinetobacter H-NS-like protein was able to restore H-NS dependent phenotypes to a *hns* null mutant strain whereas the Psychrobacter protein was only able to restore the wild type phenotype when the temperature was below 30°C; the N-terminal domain was found to have lower thermal stability with respect to the *E. coli* and *Acinetobacter* proteins (Tendeng et al., 2003).

Interestingly the KorB protein of plasmid pKM101 contains a tandemly duplicated amino acid sequence that corresponds to two copies of the H-NS C-terminal domain (Dorman et al., 1999). It is unclear at this stage whether this protein represents a conserved nucleic acid binding configuration for H-NS-like proteins.

1.5 REGULATION OF H-NS

H-NS negatively autoregulates its own transcription (Dersch et al., 1993; Falconi et al., 1993; Free and Dorman, 1995; Dersch et al., 1993; Falconi et al., 1993; Free and Dorman, 1995; Ueguchi et al., 1993). The genes *hns* and *stpA* were shown to exert parallel autogenous control and cross-regulation, with over expression of StpA resulting in repression of *hns* transcription and vice versa (Sonden and Uhlin, 1996; Zhang et al., 1996). The number of H-NS molecules present in the cell has been studied using a variety of different techniques. Western blot analysis showed upto 22000 molecules of

H-NS per cell during exponential phase growth but this population dropped to 7000 molecules upon entry into stationary phase (Azam et al., 1999). Northern blot analysis showed a correlation between the level of H-NS expression and DNA synthesis leading to a model where *hns* autoregulation results in a relatively constant ratio of H-NS to cellular DNA. H-NS mRNA transcript levels were shown to reduce at the onset of stationary phase in line with the reduced level of DNA synthesis. (Free and Dorman, 1995).

Several 'specific' H-NS binding sites have been reported in the region of the *hns* promoter. Two high affinity H-NS binding sites were identified flanking a static bend located 150 base pairs upstream of the *hns* promoter. A third lower affinity site was found overlapping the -35 element of the promoter (Falconi et al., 1993). Removal of the two high affinity binding sites (bases -90 to -606 upstream of the *hns* transcriptional start site) results in a loss of transcriptional repression by H-NS. Seven binding sites for the nucleoid associated protein Fis have been identified in the *hns* promoter region, some of which overlap the H-NS binding sites. Increasing Fis concentration leads to increased activation of the *hns* promoter in a chloramphenicol acetyl transferase reporter gene assay (Falconi *et al.*, 1996).

hns gene expression is also regulated at the level of mRNA translation. DsrA is a small non-coding mRNA molecule that affects the expression of H-NS and the stationary phase and stress response sigma factor RpoS. DsrA interacts with both the 5' and 3' ends of *hns* mRNA transcripts to inhibit *hns* translation by blocking translation initiation and enhancing turnover of the *hns* mRNA. H-NS represses the expression of *dsrA* itself forming a negative feedback loop (Lease et al., 1998; Lease and Belfort, 2000).

1.6 H-NS Domain Architecture

H-NS is a 15.6 kDa protein with two functional domains separated by a short flexible linker. Various truncations and point mutants tested *in vivo* and *in vitro* first hinted at the presence of two functionally distinct domains. Experiments using C-terminal truncations of this protein showed that the N-terminal 90 residues are sufficient for high order oligomerisation in a concentration-dependent manner comparable to that of the full length protein. However, the N-terminal 65 residues can only interact to form a discrete dimer (Smyth et al., 2000; Badaut et al., 2002; Bertin et al., 1999; Dorman et al., 1999; Tendeng and Bertin, 2003). Truncation of the N-terminal eleven amino acids of the protein gives rise to a species capable of forming homodimers but is again

unable to form larger oligomeric species (Esposito et al., 2002). Mutations within the N-terminal domain between residues 12 and 30 have been shown to disrupt the self association of H-NS. In particular mutation of leucine 30 to proline or aspartate disrupts dimer formation (Dorman et al., 1999; Ueguchi et al., 1997). The leucine is part of a heptad repeat responsible for coiled-coil dimerisation (Esposito et al., 2002; Bloch et al., 2003). Mutations to basic amino acids at the N-terminal end of the oligomerisation domain have been reported to disrupt DNA binding. Mutation of alanine 18 to glutamate created a dominant negative phenotype resulting in the reduced repression of the *proU* and *fimB* promoters (Donato and Kawula, 1999). A H-NS R11E/R14A double mutant does not preferentially bind to curved DNA sequences (Bloch et al., 2003) unlike wild type H-NS. These arginine residues are thought to be involved in DNA binding although DNA gel shift experiments using the N-terminal domain of H-NS do not show any DNA binding activity (Shindo et al., 1995; Badaut et al., 2002).

Mutations within the C-terminal domain (residues 90 to 137) of the protein produce dominant negative H-NS proteins *in vivo*. These mutants reduce the affinity of H-NS for DNA (Spurio et al., 1997; Ueguchi et al., 1996; Williams and Rimsky, 1997). A ^{1}H 1D NMR spectra of the full length H-NS protein has the same characteristic cross peaks as the spectra observed for amino acids 90-137 alone; the C-terminal domain is free to move independently in solution from the N-terminal domain. The protons from the N-terminal domain have very few observable cross peaks suggesting that this domain is too big to be visible by NMR under the conditions tested (Smyth et al., 2000). This confirmed the previous observation that the NOE connectivities assigned to the C-terminal domain (A90 to Q137 of *E. coli* H-NS) are observed in an overlay of the NOESY and COSY spectra of the full length H-NS protein but most of the C^{α}H-NH COSY crosspeaks for the N-terminal domain were not visible (Shindo et al., 1995). The difference in visibility of the two domains of H-NS in an NMR experiment indicates the presence of a flexible linker region between residues 65 and 90. In limited proteolysis experiments cleavage of the full length H-NS protein preferentially occured between residues 83 and 90 showing that these residues are less structured and thus more accessible to protease digestion (Cusick and Belfort, 1998). The linker between the two domains of H-NS is thought to comprise amino acids 78 to 90 as these are less well conserved relative to the the rest of the protein (Dorman et al., 1999) but the exact limits of the oligomerisation domain are yet to be determined experimentally.

Secondary structure predictions based on the amino acid sequence of H-NS predict the presence of three α-helices: α-helix 1 spanning from Glu3 to Glu20, α-helix 2 from Leu23 to Ala67 and α-helix 3 between Pro72 and Ala81 (Figure 1-6). When these

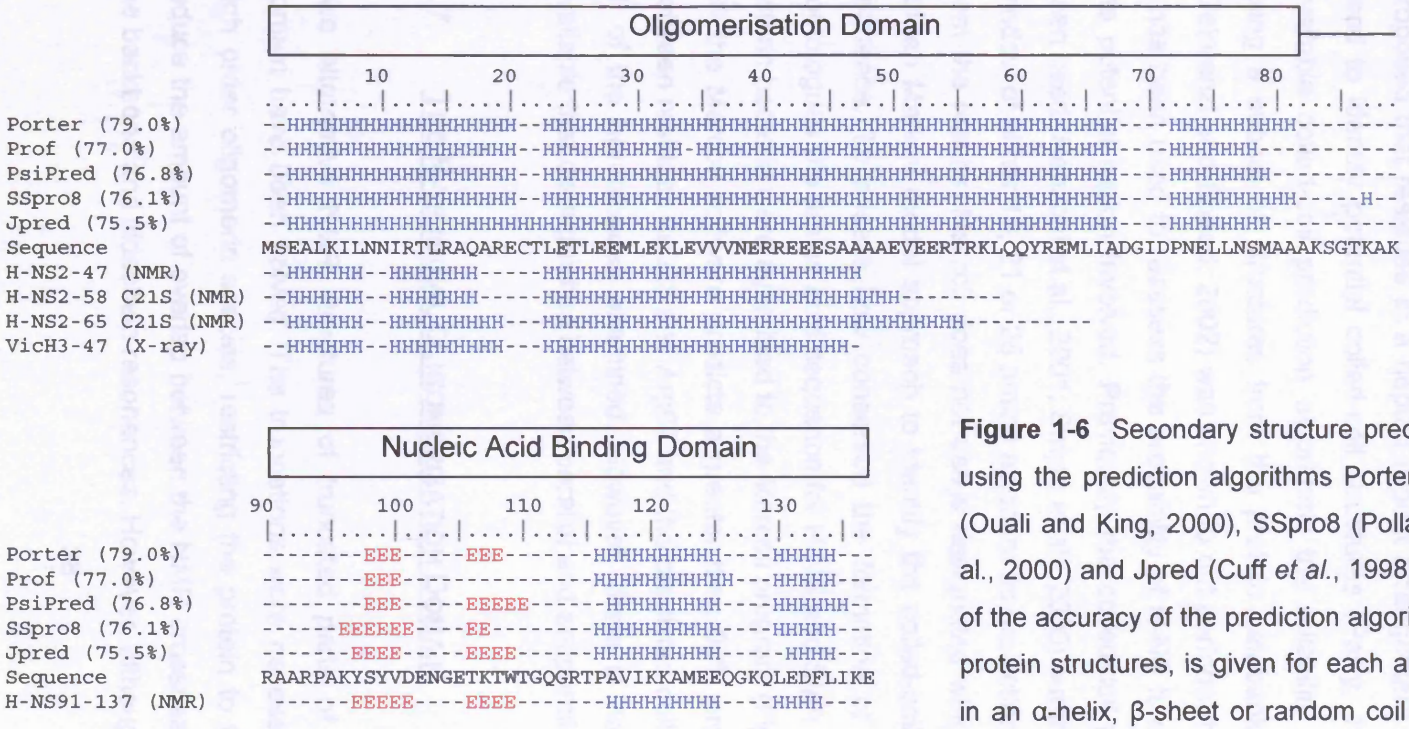


Figure 1-6 Secondary structure prediction for *S. typhimurium* H-NS protein using the prediction algorithms Porter (Pollastri and McLysaght, 2005), Prof (Ouali and King, 2000), SSpro8 (Pollastri et al., 2002), PsiPred (McGuffin et al., 2000) and Jpred (Cuff *et al.*, 1998). The Q_3 score, which gives a measure of the accuracy of the prediction algorithm when tested on a dataset of known protein structures, is given for each algorithm. An amino acid predicted to be in an α-helix, β-sheet or random coil is indicated by H, E or – respectively. The amino acid sequence for *S. typhimurium* H-NS is shown below the secondary structure predictions. For comparison the experimentally determined secondary structure assignments for H-$NS_{2\text{-}47}$ (Bloch et al., 2003), H-$NS_{2\text{-}58}$ C21S (Esposito et al., 2002), H-$NS_{2\text{-}65}$ C21S (Renzoni et al., 2001), VicH (Cerdan et al., 2003) and H-$NS_{91\text{-}137}$ (Shindo et al., 1995) are shown below.

predictions are compared with the experimentally determined secondary structure determined for H-NS_{2-47} (Bloch et al., 2003), H-NS_{2-58} C21S (Esposito et al., 2002) and H-NS_{2-65} (Renzoni et al., 2001) it appears that the predicted α-helix 1 is actually two short α-helices.

1.6.1 Coiled-Coil Prediction

A variety of coiled-coil prediction tools have been developed since 1982 when it was first proposed that residues in a heptad repeat arrangement in a protein sequence could be used to identify potential coiled-coil structures (Parry, 1982). In a recent review of the available coiled-coil prediction algorithms the reliability of these programs was tested using a subset of structures from the protein databank (Gruber *et al.*, 2006). Marcoil (Delorenzi and Speed, 2002) was shown to out perform the other prediction algorithms so it has been used to reassess the probability of H-NS forming a coiled-coil and to identify the potential region involved. Previously the coiled-coil prediction algorithm COILS has been used (Renzoni et al., 2001; Smyth et al., 2000) but this program requires the use of a window of either 14, 21 or 28 amino acids for its prediction resulting in varying conclusions from the results. Marcoil does not use a designated window size but instead relies on a hidden Markov model approach to identify the coiled-coil region in a protein amino acid sequence. To assess how conserved the formation of a coiled-coil is amongst H-NS homologues the amino acid sequence for H-NS and StpA proteins from several species of enteric bacteria were submitted to the Marcoil program (Figure 1-7). For *S. typhimurium* H-NS the Marcoil software predicts a greater than 90% probability of a coiled-coil forming between residues Leu23 and Arg62 and the predicted coiled-coil for this region is found in all of the homologues examined. However, there is no coiled-coil prediction software available that distinguishes between parallel and antiparallel coiled-coils.

1.7 THE N-TERMINAL OLIGOMERISATION DOMAIN

Two alternative NMR structures of truncated parts of the N-terminal oligomerisation domain have been solved. The truncations were necessary to prevent the formation of high order oligomeric species, restricting the protein to forming discrete dimers, and to reduce the amount of overlap between the NMR crosspeaks facilitating the assignment of the backbone and sidechain resonances. However, although the two structures share the

Figure 1-7 Marcoil coiled-coil prediction for several H-NS homologues showing the predicted probability of a coiled-coil forming for the H-NS amino acid sequence. H-NS from *S. typhimurium*, *B. aphidicola*, *H. influenzae* as well as StpA from *S. typhimurium* and *E. coli* were submitted to the Marcoil prediction algorithm using the 9FAM matrix (Delorenzi and Speed, 2002)

same secondary structure elements the relative orientation of the α-helices with respect to each other is different between the two structures. The N-terminal 58 residues from *S. typhimurium* H-NS form a homodimeric protein with a left handed parallel coiled-coil structure formed from the third helix of the two polypeptide chains (Figure 1-8). Each polypeptide has three α-helical secondary structure elements; two short α-helices, H1 (residues 3-8) and H2 (residues 11-17), and a longer third α-helix, H3 (residues 23-50). α-helices H1 and H2 fold back onto the central coiled-coil with this arrangement being stabilised by the hydrophobic residues within the coiled-coil region and a intermolecular salt bridge between Arg14 from each protomer with Glu24 and Glu27 from the other polypeptide chain (Esposito et al., 2002). In this structure a single point mutation, changing Cys21 to serine, was introduced. This mutation has been shown not to affect the folding of H-NS or the ability to form high order oligomers (Smyth et al., 2000).

An alternative NMR structure for the oligomerisation domain, from *E. coli* H-NS, consisting of the first 47 residues has also been published (Figure 1-9). This structure contains the same secondary structure elements except the third helix has been truncated further to residue 47. The relative orientation of the helices of the two protomers is different as the third helices pack together in an antiparallel coiled-coil manner (Bloch et al., 2003). There is additionally a X-ray structure of VicH (a homologue of H-NS from *Vibrio cholerae*) where the oligomerisation domain has been truncated by proteolysis to give a structure for the residues 3-50 of one protomer and 3-47 of the second polypeptide bound as a homodimer (Cerdan et al., 2003). This structure is in agreement with the NMR structure of H-NS_{1-47} in terms of its relative orientation of the secondary structure elements of the protein (Figure 1-10).

There are several possible reasons for these two different models of the H-NS N-terminal domain. Firstly significant difficulty is encountered in applying NMR techniques to solve the protein structure of this domain. The backbone amide NH cross peaks tend to cluster around 8 parts per million for alpha helical proteins leading to signal overlap in the ^{15}N-edited spectra. A number of clusters of identical amino acids in the primary sequence of H-NS give rise to further overlap in the NMR spectra. There is also considerable difficulty in assigning NOE distance constraints to this NMR structure as symmetry related degeneracy leads to ambiguity between inter- and intra-monomer NOE's. These factors could result in an incorrect structure of the protein being determined. Secondly the structures of these domains are of different truncated forms of the oligomerisation domain and may not be representative of the structure found in the wild type protein. The removal

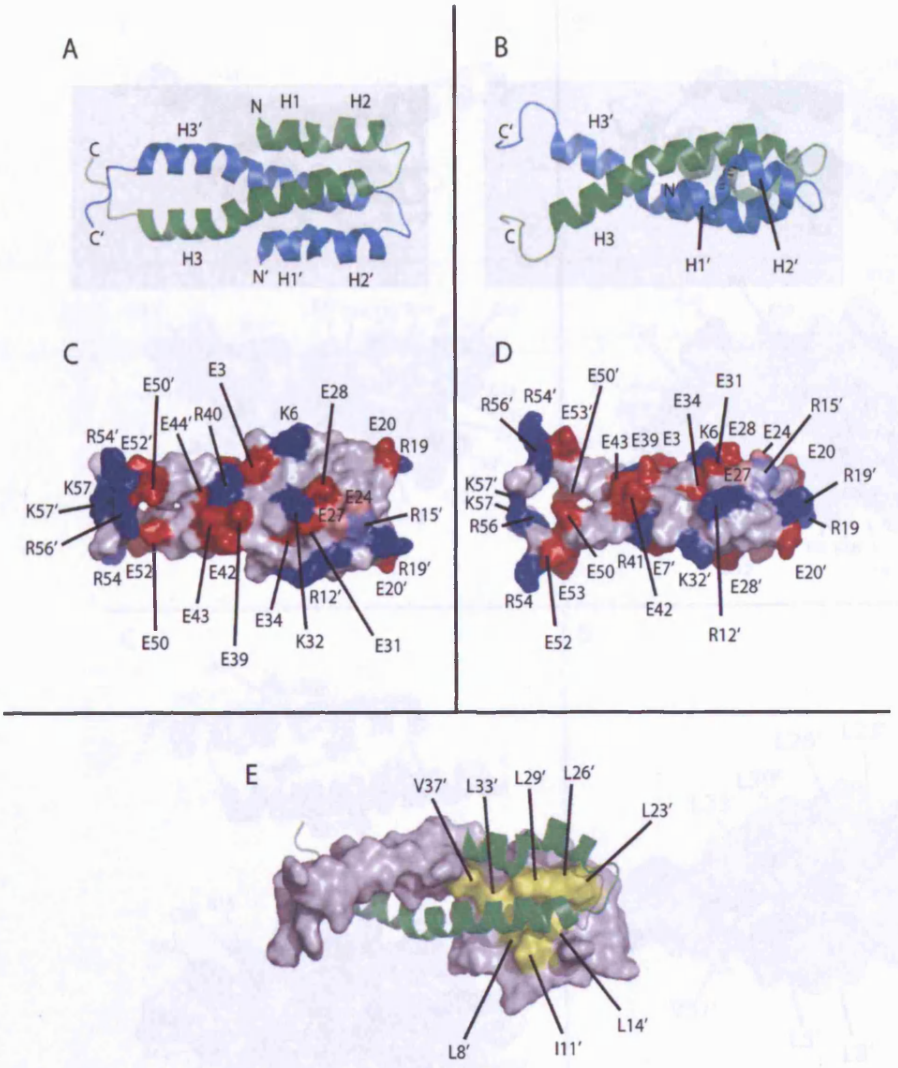


Figure 1-8. The NMR structure of *Salmonella typhimurium* H-$NS_{2\text{-}58}$ C21S. Figures A and B show cartoon depictions of the homodimer orientated at 90° rotation with respect to each other. The two protomers are coloured green and cyan respectively. The three α-helices H1, H2 and H3 or H1', H2' and H3' from each protomer are labelled. Figure C and D show the electrostatic surface of the protein with the positive (Arg and Lys) residues coloured blue and negative (Glu) residues coloured red. Glu24 and Glu27 (pale blue) and Arg15 (pink) from the other protomer form a salt bridge stabilising the position of the second helix against the core coiled-coil region. The coiled-coil packing interactions are shown in Figure E. One protomer is depicted as a ribbon structure (green) whilst the other protomer is shown with van der Waals surface representation (grey). The hydrophobic residues that stabilise the coiled-coil interaction are shown in yellow. These figures were created using Pymol (Delano, 2006).

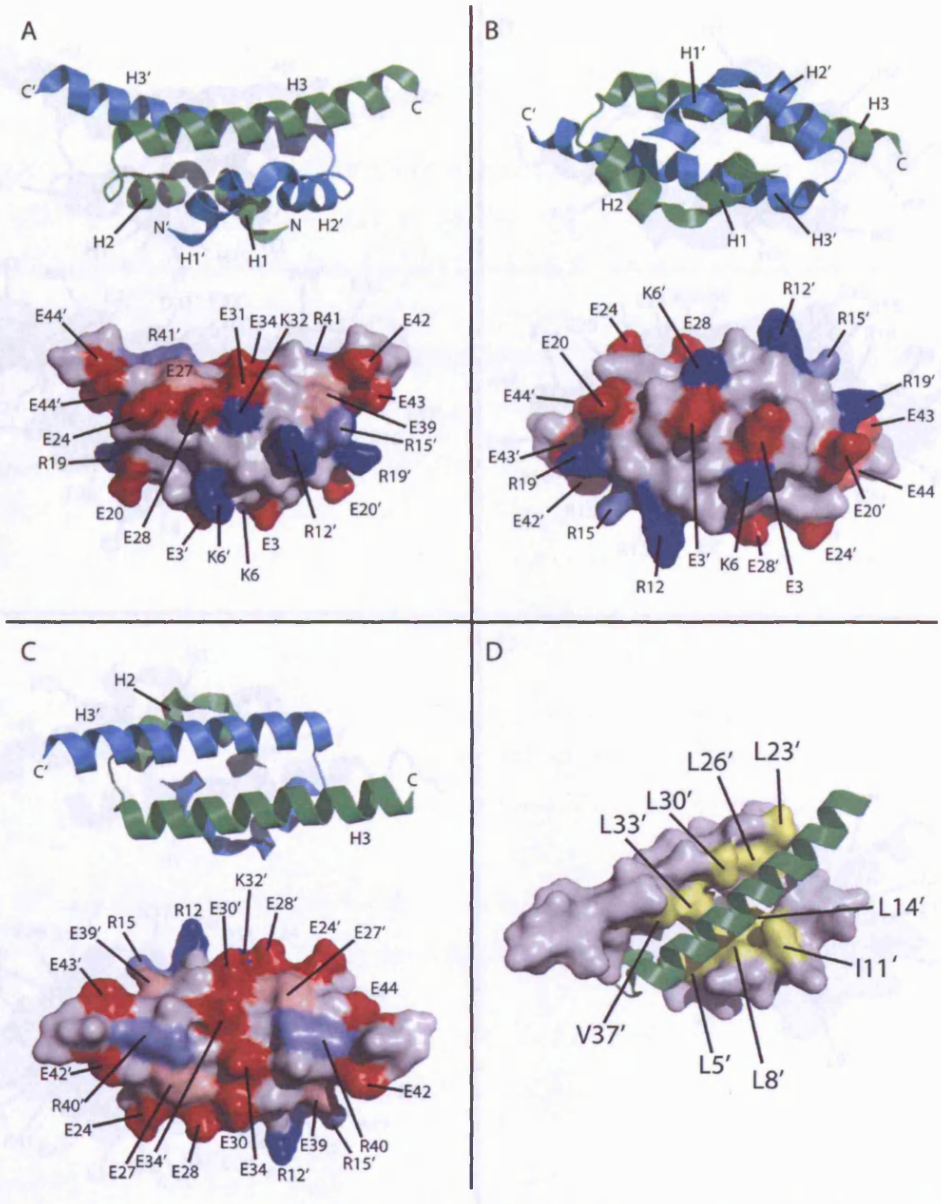


Figure 1-9 The structure of *E. coli* H-NS_{2-47} solved by NMR (Bloch et al., 2003). Figures A, B and C show a cartoon representation of the homodimer showing the two protomer chains in green and cyan respectively. Below each cartoon diagram is a van der waals surface representation of the same structure depicting the electrostatic surface of the protein where the positive (Arg and Lys) residues have been coloured blue and negative (Glu and Asp) residues coloured red. The charged residues that form intramolecular salt bridges (Glu27-Arg41', Glu27'-Arg41, Arg15-Glu39' and Arg15'-Glu39) are coloured pink (Glu) and pale blue (Arg) respectively. Figure E shows the coiled-coil packing interactions. One protomer is depicted as a ribbon structure (green) whilst the other protomer is shown as a van der Waals surface representation (grey). The hydrophobic residues that stabilise the coiled-coil interaction are shown in yellow. These figures were created using Pymol (Delano, 2006).

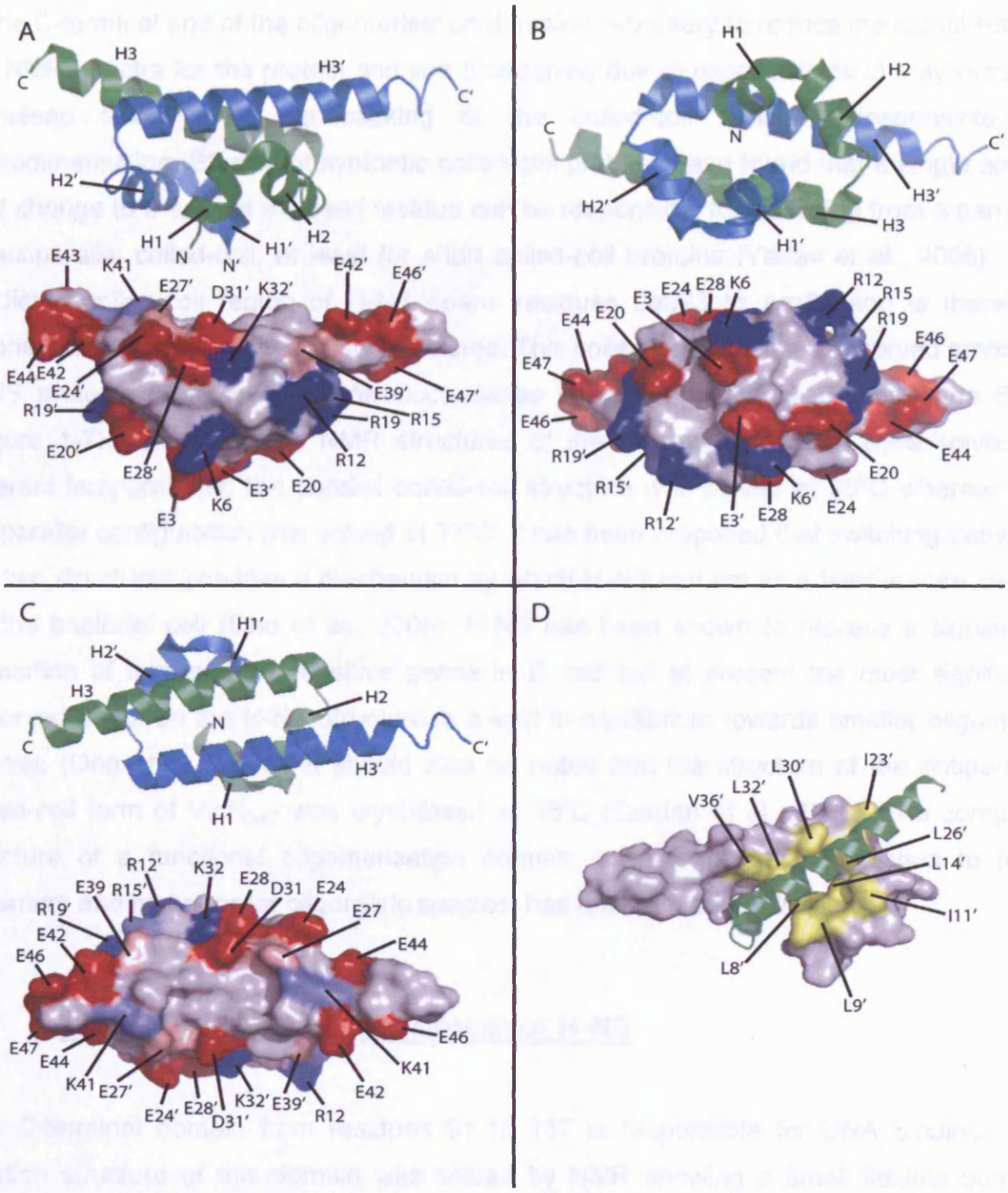
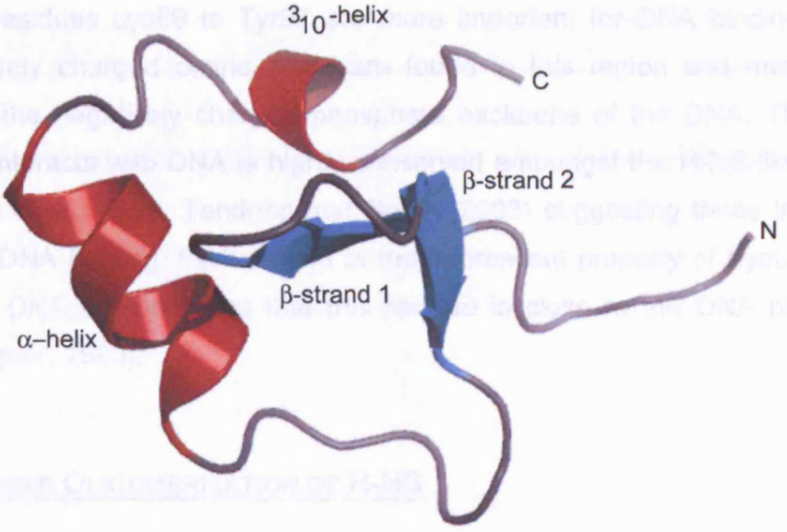


Figure 1-10 The structure of *V. cholerae* VicH homodimer solved by Xray crystallography (Cerdan et al., 2003). Figures A, B and C show a cartoon representation of the homodimer showing the two protomer chains ($VicH_{3-47}$) green and ($VicH_{3-50}$) cyan respectively. Below each cartoon diagram is a van der waals surface representation of the same structure depicting the electrostatic surface of the protein where the positive (Arg and Lys) residues have been coloured blue and negative (Glu and Asp) residues coloured red. The charged residues that form intramolecular salt bridges (Glu27-Arg41', and Glu39-Arg15') are coloured pink (Glu) and pale blue (Arg) respectively. Figure E shows the coiled-coil packing interactions. One protomer is depicted as a ribbon structure (green) whilst the other protomer is shown as a van der Waals surface representation (grey). The hydrophobic residues that stabilise the coiled-coil interaction are shown in yellow. These figures were created using Pymol (Delano, 2006).

of the C-terminal end of the oligomerisation domain (necessary to reduce the complexity of the NMR spectra for the protein and line broadening due to oligomerisation) may result in increased flexibility in the packing of the coiled-coil residues, responsible for homodimerisation. Studies of synthetic coiled-coil proteins have found that a single amino acid change to a solvent exposed residue can be responsible for switching from a parrallel to antiparallel coiled-coil, at least for short coiled-coil proteins (Yadav *et al.*, 2006). The predicted coiled-coil region of H-NS spans residues Leu23 to Arg62 and is therefore incomplete in both published NMR structures. This coiled-coil region is conserved amongst H-NS proteins from several *enterobacteriaciae* species and the H-NS paralogue StpA (Figure 1-7). The published NMR structures of the N-terminal domain were solved at different temperatures; the parallel coiled-coil structure was solved at 25°C whereas the antiparallel configuration was solved at 37°C. It has been proposed that switching between the two structures provides a mechanism by which H-NS can act as a temperature switch for the bacterial cell (Ono et al., 2005). H-NS has been shown to repress a significant proportion of temperature sensitive genes in *E. coli* but at present the most significant observed affect on the H-NS structure is a shift in equilibrium towards smaller oligomeric species (Ono et al., 2005). It should also be noted that the structure of the antiparallel coiled-coil form of $VicH_{3-47}$ was crystallised at 18°C (Cerdan et al., 2003). The complete structure of a functional oligomerisation domain, capable of self association to form tetramers and higher order oligomeric species, has not been determined.

1.8 THE NUCLEIC ACID BINDING DOMAIN OF H-NS

The C-terminal domain from residues 91 to 137 is responsible for DNA binding. The solution structure of this domain was solved by NMR showing a small flexible domain formed from two short β-sheets, a single α-helix and a 3_{10}-helix (Figure 1-11). Relatively few NOE's were observed for the amide protons in this protein suggesting a fairly mobile structure for this domain. This is reflected in the root mean squared deviation (RMSD) determined for the backbone heavy atoms for the 16 lowest energy NMR structures of 1.59 Å (Shindo et al., 1995). DNA binding studies by NMR using the extended $H\text{-}NS_{60-137}$ construct identified two loop regions in the H-NS protein that are able to interact with the DNA. Chemical shifts were observed, upon addition of DNA, to cross peaks corresponding to amino acids between Ala81 and Lys96 and from Thr110 and Ala117 (Shindo et al., 1999). These two regions are reflected in the clustering of loss of function point mutants




```
                                100       110       120       130
                       ....|....|....|....|....|....|....|....|....|..
H-NS E. coli           AQRPAKYSYVDENGETKTWTGQGRTPAVIKKAMDEQGKSLDDFLIKQ
```

Figure 1-11 Cartoon depiction of H-NS_{91-137} structure solved by NMR (Shindo et al., 1995). The amino acid sequence for this construct is shown with the helical regions coloured red and the β-strands coloured cyan. Underlined in the amino acid sequence are the regions involved in DNA binding (Shindo et al., 1995; Shindo et al., 1999). This figure was created using Pymol (Delano, 2006).

that have been identified between amino acids Arg90 to Tyr97 and Thr110 to Ile119 (Ueguchi et al., 1996; Williams and Rimsky, 1997). The first region is the flexible linker between the oligomerisation domain and the DNA binding domain. Residues Ala81 to Ala88 are less well conserved amoungst H-NS homologues than residues Lys89 to Tyr97 suggesting that residues Lys89 to Tyr97 are more important for DNA binding. Several conserved positively charged amino acids are found in this region and may facilitate interactions with the negatively charged phosphate backbone of the DNA. The second loop region that interacts with DNA is highly conserved amoungst the H-NS-like family of proteins (Dorman et al., 1999; Tendeng and Bertin, 2003) suggesting these interactions are essential for DNA binding. Perturbation of the fluorescent property of tryptophan 109 upon addition of DNA demonstrates that this residue is close to the DNA binding site (Tippner and Wagner, 1995).

1.9 HIGH ORDER OLIGOMERISATION OF H-NS

The N-terminal domain of H-NS mediates the formation of homodimers and high order oligomeric species. This self-association is critical for H-NS function as a transcriptional repressor and is necessary for preferential binding to curved DNA sequences (Badaut et al., 2002; Stella et al., 2005; Ueguchi et al., 1996; Ueguchi et al., 1997; Williams and Rimsky, 1997). Truncations and point mutations have been widely studied to assess the self association of H-NS. Truncation of the N-terminal end of the oligomerisation domain restricts the protein to the formation of a discrete homodimer. H-NS_{13-90} (Esposito et al., 2002), H-NS_{61-137} (Shindo et al., 1995) both form stable homodimers. Deletion of the C-terminal end of the oligomerisation domain (residues 66-90) also leads to the restriction of H-NS to a homodimeric species, as observed for H-NS_{1-65} (Badaut et al., 2002; Esposito et al., 2002). Single amino acid substitution mutations to introduce a proline residue into each of the first three α-helices of H-NS provides evidence for the importance of each helix for oligomerisation and H-NS function. The first helix is not critical for high order oligomerisation whereas disruption of the second helix, by a Q17P point mutation, restricts H-NS to a dimeric structure (Esposito et al., 2002). Substitution of Leu26 for proline as well as the double mutation E53G/T55P prevent high order oligomerisation but is not sufficient to block dimerisation (Badaut et al., 2002). These results show that high order oligomeric species of H-NS (complexes larger than homodimers) will only form if the full oligomerisation domain is intact. Disruption to either the N-terminal end or the C-terminal end of the oligomerisation domain prevents higher order oligomerisation.

Two alternative models have been proposed to explain the higher order oligomerisation of H-NS based on either the parallel coiled-coil H-NS homodimer or the antiparallel coiled-coil H-NS homodimer. Figure 1-12 depicts the two different models for the polymerisation of H-NS homodimers. The first model is called the head to tail model. In this model the N-terminal residues (the head) of the H-NS homodimer bind to the residues between 65 and 90 (the tail) to form a tetramer or larger oligomeric species. This would leave the nucleic acid binding domains free to access DNA on either side of the polymerised protein structure (Esposito et al., 2002). The relative orientation of the H-NS homodimers is not known. The H-NS homodimers could interact in an end on manner, as depicted in Figure 1-12, possibly with some overlap of the main third α-helices of the H-NS homodimers or the H-NS homodimers could bind antiparallel to each other, with much greater overlap of the central coiled-coil region of the protein.

An alternative model for the formation of high order oligomeric species of H-NS suggests that separate head to head and tail to tail interactions are responsible for the polymerisation of the protein. In this model anti-parallel coiled-coil H-NS homodimers, based on the structure of H-NS_{2-47} (Bloch et al., 2003), interact via head to head and tail to tail interactions so two distinct interactions are required for the formation of tetramers or larger oligomeric species.

It has also been claimed, based on substitution and deletion mutants, that amino acids within the nucleic acid binding domain are involved in high order oligomerisation (Spurio et al., 1997; Stella et al., 2005). The experiments used in this case did not measure the high order oligomerisation directly, but instead measured the repression of a reporter gene *in vivo* so a disruption to the DNA binding interaction of H-NS may also be able to explain the reduced level of reporter gene repression in the H-NS P116A and H-NS P116Δ mutants. In the absence of the DNA binding domain H-NS_{1-90} is still able to form high order oligomeric species comparable to those observed with the full length protein (Lease et al., 1998; Lease and Belfort, 2000; Smyth et al., 2000).

1.10 H-NS DNA Binding

H-NS has a preference for AT-rich DNA sequences *in vitro* binding preferentially to planar curved DNA, compared to twisted curved DNA, moderately curved DNA or non curved

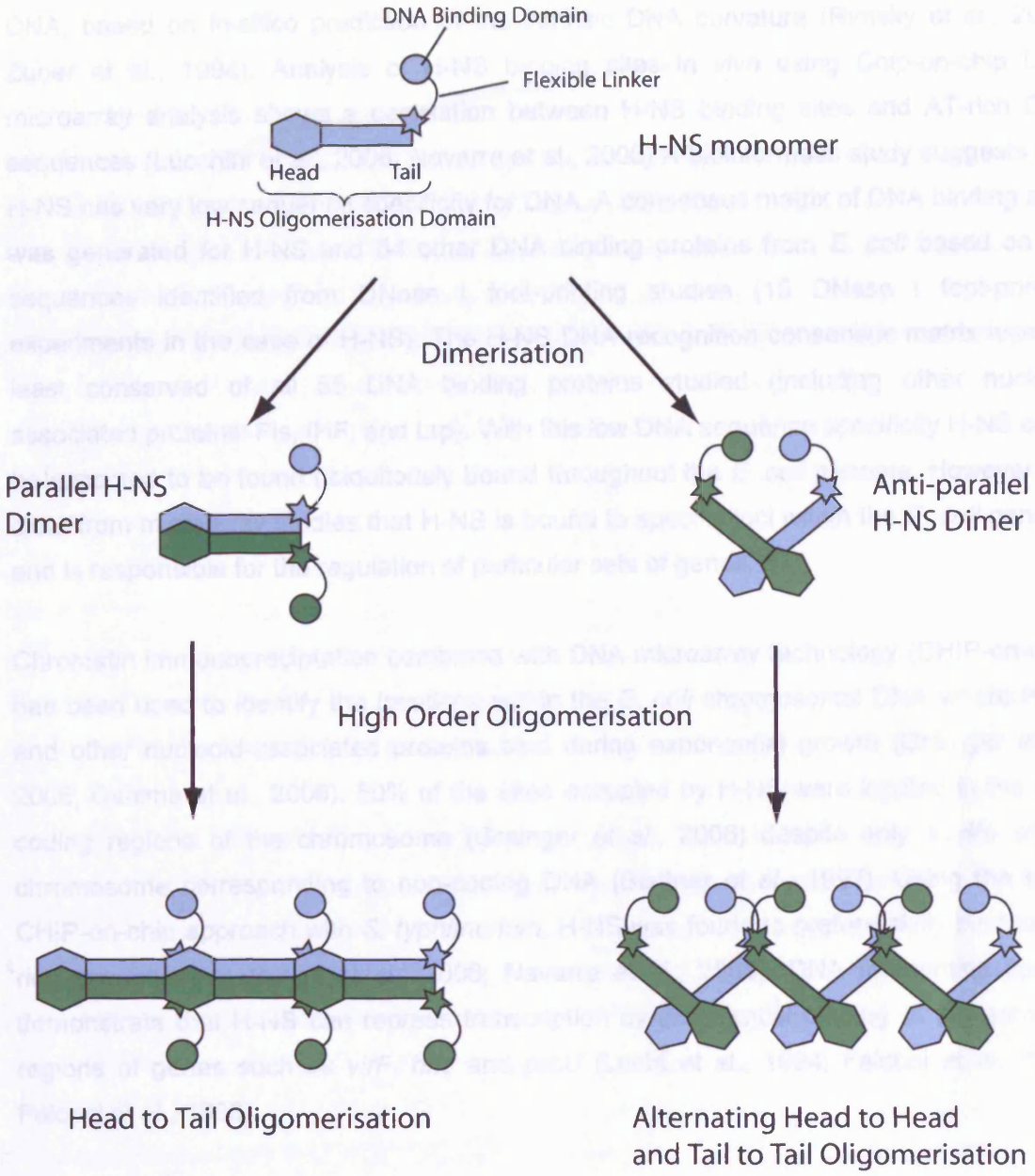


Figure 1-12. Alternative models for the high order oligomerisation of H-NS. Parallel coiled-coil H-NS homodimers combine via head-to-tail interactions to form tetramers and higher order oligomeric complexes. Anti-parallel coiled-coil H-NS homodimers combine via tail-to-tail and head-to-head interactions to form polymers of H-NS .

DNA, based on in-silico prediction of the intrinsic DNA curvature (Rimsky et al., 2001; Zuber et al., 1994). Analysis of H-NS binding sites in *vivo* using Chip-on-chip DNA microarray analysis shows a correlation between H-NS binding sites and AT-rich DNA sequences (Lucchini et al., 2006; Navarre et al., 2006) A bioinformatic study suggests that H-NS has very low sequence specificity for DNA. A consensus matrix of DNA binding sites was generated for H-NS and 54 other DNA binding proteins from *E. coli* based on the sequences identified from DNase I foot-printing studies (15 DNase I foot-printing experiments in the case of H-NS). The H-NS DNA recognition consensus matrix was the least conserved of all 55 DNA binding proteins studied (including other nucleoid associated proteins: Fis, IHF, and Lrp). With this low DNA sequence specificity H-NS could be expected to be found ubiquitously bound throughout the *E. coli* genome. However it is clear from microarray studies that H-NS is bound to specific loci within the *E. coli* genome and is responsible for the regulation of particular sets of genes.

Chromatin immunopreciptation combined with DNA microarray technology (CHIP-on-chip) has been used to identify the locations within the *E. coli* chromosomal DNA where H-NS and other nucleoid-associated proteins bind during exponential growth (Grainger et al., 2006; Oshima et al., 2006). 50% of the sites occupied by H-NS were located in the non-coding regions of the chromosome (Grainger *et al.*, 2006) despite only 11.4% of the chromosome corresponding to non-coding DNA (Blattner *et al.*, 1997). Using the same CHIP-on-chip approach with *S. typhimurium,* H-NS was found to preferentially bind to AT-rich sequences (Lucchini et al., 2006; Navarre *et* al., 2006). DNA footprinting studies demonstrate that H-NS can repress transcription by preferential binding to the promoter regions of genes such as *virF*, *hns* and *proU* (Lucht et al., 1994; Falconi et al., 1993; Falconi et al., 1998).

1.11 H-NS TARGET GENES

H-NS plays an important role in mediating the response to changes in environment such as the adaptation to changes in temperature, osmolarity, pH, and oxygen availability. DNA microarray analysis of mRNA levels in wild type *E. coli* and an *E. coli* strain with a *hns* null mutation showed a significant difference in mRNA expression level for about 250 genes (Hommais et al., 2001). 30% of these H-NS controlled genes are involved in the cell envelope and cellular processes, 8% in intermediary metabolism and 21% in information pathways. Further microarray experiments investigating the effect of temperature on gene

expression, in both wild type *E. coli* and a *hns* null mutant strain, identified 531 genes (12% of the *E. coli* genome) whose mRNA expression level increased by at least three fold in response to temperature. From these 531 genes, 408 genes (77% of the temperature affected genes) were shown to be controlled, either directly or indirectly, by H-NS (Ono et al., 2005). Size exclusion chromatography experiments show a decrease in the average size of the H-NS oligomeric complex as the temperature is increased from 17.5°C to 45°C and a decrease in the binding affinity of H-NS for DNA was observed over the same temperature range (Ono et al., 2005). Temperature induced changes in the topology of the DNA (Drlica, 1992; Karem and Foster, 1993) and the formation of heteromeric complexes between H-NS and Hha/YmoA family proteins (Nieto et al., 2002) have also been proposed as possible mechanisms for the modulation of H-NS-mediated gene repression in response to temperature.

Many of the genes that are under the control of H-NS are important for bacterial virulence in a mammalian host. In the soil, *E. coli* grows anaerobically at varying osmolarities. Upon entering the mammalian gastrointestinal tract *E. coli* adapts to the high osmolarity anaerobic environment. The adaptation of *E. coli* to hyperosmolarity environments requires an initial rapid uptake of K^+ ions to prevent dehydration prior to the accumulation of the osmoprotectants glutamate and trehalose. The osmoprotectants glycine betaine and proline can be taken up from the surrounding environment via two osmotically regulated permeases ProP and ProU. The repression of the *proU* operon by H-NS involves the binding of H-NS to a regulatory element located 270 bases downstream of the transcription start site (Owen-Hughes et al., 1992; Jordi and Higgins, 2000). In addition two identical DNA sequences at positions +25 and +130 have been identified as high affinity H-NS binding sites (Bouffartigues et al., 2007). A bioinformatic study searched for similar sequences to the 5'-TCGATATATT-3' sequence idenitified at the *proU* operon within H-NS DNase I footprints from six genes known to be repressed by H-NS to build up a consensus H-NS binding sequence. Based on this consensus sequence H-NS binding sites in the *E. coli* chromosome were predicted (Lang et al., 2007) and compared with H-NS binding sites identified by Chip-on-chip microarray analysis (Grainger et al., 2006). The consensus sequence was identified in 59 of the 99 highest scoring sites from the DNA microarray analysis. Whilst the consensus sequence was identified in a statistically significant number of the high affinity H-NS binding sites the correlation between this consensus sequence and all known H-NS binding sites shows little agreement (Lang et al., 2007) so whilst the consensus sequence may have an important role in directing H-NS binding at a

subset of H-NS repressed genes it cannot explain the affinity of H-NS for all of its binding sites.

H-NS is involved in the response to acid tolerance by repressing the *ompR* acid induced response regulator under normal conditions. In a *hns* mutant a constitutive acid tolerance response is observed (Bang et al., 2002) along with greater resistance to low pH relative to wild type *E. coli* (Bertin et al., 2001). In addition, H-NS represses the *flhDC* operon which encodes proteins that affect cell motility, chemotaxis and the differentiation of *enterobacteria* into swarming cells.

1.12 Mode of Action for Transcription Repression by H-NS

H-NS does not recognise a specific sequence of DNA but has enhanced affinity for planar curved DNA structures. The inherent curvature of AT-rich sequences, often found at promoter regions, contributes to the preferential binding of H-NS to the non-coding regions of the bacterial chromosome but is not sufficient to explain H-NS specificity for its target promoters (Jordi et al., 1997). Curved regions of the DNA can be found at most promoters including those for genes not repressed by H-NS (Ohyama, 2001).

Bridging between two regions of DNA is a common feature of H-NS mediated gene repression identified in the best characterised promoters studied such as the *virF, hns*, *proU* and the *rrnB P1* promoter. The binding of H-NS to the *virF* promoter region is illustrated in Figure 1-13. DNA footprinting studies have identified two H-NS binding regions within a short region of intrinsically curved DNA at the *virF* promoter of *Shigella flexneri*. One H-NS binding site overlaps the *virF* promoter whilst the second site is located approximately 250 base pairs upstream of the promoter. H-NS is thought to bridge between the two regions of DNA, inducing the formation of a hairpin-like structure that blocks access of RNA polymerase to the promoter. DNA bridging between the two regions of H-NS binding is essential for the temperature sensitive repression of the *virF* promoter. A change in temperature from 25°C to 37°C reduces the degree of curvature between the cis-acting H-NS binding sites and also displaces the curve centre (Prosseda et al., 2004). Insertion of a half turn of DNA helix between the two H-NS binding regions also disrupts the H-NS mediated repression of the *virF* promoter *in vivo*. At the *virF* promoter, H-NS competes with another nucleoid associated protein called Fis, which binds to other sites in

Figure 1-13 The repression of the *virF* promoter in *Shigella flexneri* by H-NS. A. The temperature induced change in DNA curvature and the affect on DNA bridging by H-NS. At 25°C H-NS is able to bind to both the promoter and upstream (-250) recognition site creating a hairpin-like structure. An increase in temperature alters the curvature of the DNA so that H-NS can no longer bridge the two recognition sites. B. The antagonistic effect of Fis on H-NS repression of the *virF* promoter. Fis binds to a site between the two H-NS binding regions, within the hairpin, changing the curvature of the DNA so that H-NS can no longer bridge between its two binding sites (Prosseda et al., 2004).

the DNA upstream of the *virF* promoter and alters the conformation of the DNA so that H-NS is no longer able to bridge between the two H-NS binding regions. Fis bends DNA by 50° to 90° (Pan et al., 1996) and will therefore change the relative positions of the H-NS binding regions.

At the *rrnB* P1 promoter atomic force microscopy studies demonstrate the trapping of RNA polymerase in an open initiation complex at the promoter by H-NS (Dame et al., 2002). At this promoter two H-NS binding sites were again identified but this time neither site overlaps the promoter so RNA polymerase is able to bind and initiate transcription. H-NS blocks transcription elongation by acting as a road block to RNA polymerase. If H-NS is able to bridge between the two H-NS binding regions a hairpin-like structure is formed with the promoter region in the hairpin loop. RNA polymerase can bind to the promoter but is unable to disrupt H-NS binding and therefore cannot transcribe the rest of the *rrnB P1* gene. Fis again acts as an antagonist to H-NS binding. (Dame et al., 2002; Afflerbach et al., 1999; Lease et al., 1998; Lease and Belfort, 2000).

A different model for H-NS mediated repression of transcription has also been reported based on DNA footprinting studies using increasing concentrations of H-NS. In this model H-NS binds to a high affinity binding site. The bound H-NS then recruits other H-NS molecules to this region and the H-NS polymerises along the DNA to create tracts of DNA that are coated by H-NS so that RNA polymerase is unable to bind (Figure 1-14). Quantitative DNase I foot-printing studies of H-NS binding to the *proU* P2 promoter region identified two high affinity H-NS binding sites with the same DNA sequence (5'-TCGATATATT-3') 130 bp and 25 bp upstream of the P2 promoter (Bouffartigues et al., 2007). The DNA sequence identified was found to be susceptible to attack by potassium permanganate and has a low melting point relative to the surrounding DNA, suggesting that it encodes a single stranded or distorted DNA duplex. The susceptibility of the 5'-TCGATATATT-3' sequence for $KMnO_4$ attack was not affected by the binding of H-NS. Furthermore when the high affinity binding site was inserted into a DNA sequence with a generally low affinity for H-NS, the protein not only bound to the inserted DNA sequence but additionally to two other sites in the DNA that had not been previously recognised by H-NS. The binding affinities of H-NS for the three regions of DNA were 46 nM, 48 nM and 50 nM demonstrating the cooperative binding of H-NS. Increasing the temperature from 20°C to 37°C disrupts the cooperative binding with H-NS preferentially binding to the inserted DNA sequence at 37°C (Bouffartigues et al., 2007; Lang et al., 2007).

Figure 1-14 H-NS mediated repression of gene expression by oligomerisation of H-NS along the DNA. H-NS initially binds to a high affinity binding site. Additional H-NS molecules are then recruited to form a larger H-NS oligomer that polymerises along the DNA (Rimsky et al., 2001).

Several other proteins, as well as Fis, are able to act as antagonists or enhancers of H-NS transcription repression. The cyclic AMP receptor protein (CRP) that binds to specific DNA sequences in the presence of cyclic AMP (cAMP) antagonises the repression of *papB* by H-NS. Like Fis, CRP introduces bends in the DNA at the site of binding (Lin and Lee, 2003) changing the structure of the DNA. Leucine responsive regulatory protein (Lrp) can act as either an activator or repressor of transcription at different promoters and depending on the presence or absence of leucine (Calvo and Matthews, 1994). At the ribosomal RNA promoters Lrp and H-NS both act as transcriptional repressors and enhance the repression property of each other. This may be via a direct protein:protein interaction or by cooperative changes in DNA conformation.

The role of RNA polymerase also plays an important role in the specificity of H-NS-mediated repression for certain promoters. RNA polymerase in prokaryotes consists of a core holoenzyme of four subunits ($\alpha_2\beta\beta'$) that is responsible for the production of mRNA and a σ-subunit (sigma factor) responsible for promoter recognition. There are several different σ-subuints in *E. coli* that recognise a different, but sometimes overlapping, subset of gene promoters. The σ^{70} (or σ^{D}) sigma factor is responsible for recognition of the housekeeping genes and is responsible for transcription of most genes expressed in growing cells. The σ^{38} (or σ^{S}) sigma factor targets genes required in response to starvation or stress response. The binding of H-NS to *hdeAB* promoter DNA bound to these two sigma factors has been studied by atomic force microscopy. Transcription initiation by the RNA polymerase with the σ^{70} subunit is repressed by H-NS whilst transcription initiation when the σ^{38} subunit is present is not repressed. The σ^{70} sigma factor bends the DNA around its surface so that the DNA either side of the sigma factor binding site comes into close proximity and can be bridged by H-NS. In the case of the σ^{38} sigma factor the DNA is not bent to the same extent so H-NS is unable to bridge between the DNA regions to form a hairpin-like structure (Shin *et al.*, 2005). For example the *csgA*, *gadB/C* and *csiD* promoters in *E. coli* are repressed by H-NS in a σ^{70} dependent manner (Arnqvist et al., 1994; Marschall et al., 1998; Waterman and Small, 2003). If the *rpoD* gene encoding the RNA polymerase sigma factor, σ^{70}, is mutated then H-NS is no longer able to act as a transcription repressor at the *csgA*, *gadB/C* and *csiD* promoters because the conformation of the DNA upon binding of the remaining RNA polymerase sigma factors to these promoters is not suitable for the recruitment of H-NS.

1.13 H-NS Protects the Cell Against Horizontal Gene Transfer

Bacteria are able to acquire foreign DNA and incorporate it into their genomes but these genes need to be regulated so that their expression is not detrimental to the bacterial cell. However, if the newly acquired genes are to confer a new and useful function to their new host they must be expressed at some point (Dorman, 2007b). Analysis of H-NS binding sites in the *S. typhimurium* genomic DNA demonstrates that H-NS binds to regions with high AT content (Lucchini et al., 2006; Navarre et al., 2006). AT-rich sequences are commonly found at gene promoters including the pathogenicity islands that encode many virulence proteins in *S. typhimurium*. It is thought that the pathogenicity islands were acquired by horizontal gene transfer (Groisman and Ochman, 1997; Ochman et al., 2000) and that H-NS or H-NS-like proteins play a role in repressing gene expression of these genes. For example, H-NS is able to bind to the spv virulence gene locus in *S. typhimurium* to repress gene expression (Lucchini et al., 2006; O'Byrne and Dorman, 1994).

1.14 An Overview of StpA

StpA (suppressor of the Td^- phenotype) was first identified by its ability to supress the mutant phenotype of the *td-* gene from bacteriophage T4. StpA acts as an RNA chaperone to promote the correct splicing of the bacteriophage T4 mRNA in Td^- mutants. This 134 amino acid protein shares 57% sequence identity with the paralogous protein H-NS (Zhang and Belfort, 1992). In rich growth media transcription of the *stpA* gene is confined to a short period of logarithmic phase growth with up-regulation of the gene observed in response to osmotic stress or temperature increase (Free and Dorman, 1997). In minimal media more prolonged *stpA* expression is observed in an Lrp dependent manner (Sonden and Uhlin, 1996).

StpA, like H-NS, has an N-terminal oligomerisation domain and a C-terminal nucleic acid binding domain separated by a flexible linker. These proteins are able to form both homomeric and heteromeric complexes (Deighan et al., 2003; Williams et al., 1996) leading to the possibility of forming complexes of differing activity depending on the combination of subunits present (Dorman et al., 1999; Dorman et al., 1999; Dorman et al., 1999; Johansson and Uhlin, 1999). The StpA protein is less resistant to Lon protease

digestion than H-NS and formation of heterocomplexes with H-NS reduces the turnover of StpA within the cell (Johansson and Uhlin, 1999).

It has been suggested that StpA can act as a molecular backup for H-NS as the up-regulation of StpA expression was shown to compensate for a subset of genes normally regulated by H-NS (Zhang et al., 1996). However, DNA microarray studies following the level of mRNA transcripts involved in resistance to low pH and cell motility in *hns*, *stpA* or an *hns stpA* double mutant found that StpA was not involved in either of these two H-NS related phenotypes (Bertin et al., 2001). In uropathogenic *E. coli* 551 genes were identified as being upregulated (515 genes) or down-regulated (36 genes) in an mutant strain. However no genes were identified as having changed transcription levels in a *stpA* mutant but despite this in a *hns* and *stpA* double mutant the affected genes did not completely match those identified from the *hns* mutant only strain. (Muller *et al.*, 2006). In an *hns* mutant the level of StpA in the cell only reaches about 10% of that observed for H-NS in wild type *E. coli* so the inability of StpA to fully compensate for the loss of H-NS in a *hns* mutant may be due to this lower level of protein expression rather than a difference in specificity for particular promoters (Sonnenfield et al., 2001).

1.15 DNA Binding Properties of StpA

Both StpA and H-NS can constrain DNA supercoils (Zhang et al., 1996) *in vitro* and bind preferentially to planar curved DNA sequences (Sonnenfield et al., 2001). DNAse I footprinting treatment of *gal* promoter DNA with either H-NS or StpA bound identified the same protected region for both nucleoid associated proteins (Zhang et al., 1996). However StpA binds with four-fold higher affinity for DNA than H-NS (Sonnenfield et al., 2001).

1.16 RNA Binding Activity

StpA and H-NS are both able to bind to RNA but StpA has a ten-fold higher RNA chaperone activity compared to H-NS (Zhang et al., 1996). StpA promotes the efficient splicing of the td group I intron pre-mRNA by disrupting the tertiary structure of misfolded RNA, allowing the RNA to refold into native conformations (Mayer et al., 2002; Waldsich et al., 2002). One distinct role of StpA compared to H-NS has been suggested for the control of the OmpF porin protein expression level via the control of *micF* RNA (Deighan et al.,

2000; Delihas and Forst, 2001). The *micF* gene encodes a non-translated anti-sense RNA molecule that binds to *ompF* mRNA to inhibit translation of the *ompF* transcript. H-NS inhibits *micF* transcription resulting in increased expression of the *ompF* gene (Suzuki *et al.*, 1996) but an alternative role for StpA in the regulation of *ompF* expression has been suggested. StpA binding to *micF* RNA leads to the degradation of the *micF* RNA in a post-transcriptional regulation process (Deighan et al., 2000). The half life of *micF* RNA was measured in a *hns* mutant, after exposure to the transcription inhibitor rifampicin, was found to be 2.3 minutes whereas the *micF* RNA half life in a *stpA hns* double mutant was 12.6 minutes, showing that in the absence of StpA *micF* RNA is degraded more rapidly (Deighan et al., 2000).

1.17 Formation of Heteromeric Complexes Modulates H-NS Mediated Gene Repression

Multiple H-NS paralogues have been identified in several enteric bacteria strains and provide a mechanism for the cell to adapt to changing environmental conditions by modulating the levels of homomeric and heteromeric H-NS-like protein complexes in the cell. The interaction of H-NS with its close paralogue StpA has been well documented as discussed in section 1.10 but other H-NS-like proteins have also been identified providing further potential variations in the oligomeric complexes formed.

In *S. flexneri* the H-NS-like protein, Sfh, is present along with H-NS and StpA and has been shown to form homomeric complexes and heteromeric complexes with either H-NS or StpA by yeast two hybrid experiments. All three proteins are able to auto-repress their own gene transcription and also repress transcription at the promoters of the other two genes. The expression level of the three proteins, H-NS, StpA and Sfh, are growth phase-dependent suggesting that the variation in the levels of homomeric and heteromeric complexes formed from these proteins provides a means for the cell to adapt to the changing environment. Five *hns*-like genes have been identified in *Pseudomonas putida* showing that multiple copies of H-NS-like proteins are found in a number of different enteric bacteria (Tendeng and Bertin, 2003).

Another family of proteins that bind to the N-terminal oligomerisation domain of H-NS or StpA has been identified. The Hha/YmoA family of proteins participate in the thermoregulation and osmomodulation of gene expression in enteric bacteria by forming

heteromeric complexes with H-NS or StpA (Nieto et al., 2002; Paytubi et al., 2004). NMR and fluorescence anisotropy experiments have shown that Hha binds to the N-terminal end of a $H\text{-}NS_{1\text{-}64}$ homodimer to form a trimeric protein complex (Garcia et al., 2005; Paytubi et al., 2004) via residues at the N-terminus of H-NS (Garcia et al., 2006; Paytubi et al., 2004). Hha is thought to mimic the structure of the oligomerisation domain of H-NS to disrupt the formation of H-NS high order oligomeric complexes and to modulate the ability of H-NS to repress transcription (Madrid et al., 2007a).

1.18 CONCLUDING REMARKS

H-NS and H-NS-like proteins play an important role in controlling the adaptation of the enterobacterial cell to changes in its environment. There are several levels of regulation that control H-NS-mediated gene repression. H-NS does not appear to bind to specific DNA sequence, although a preference for AT-rich sequences seems to be important for the targeting of H-NS to non-coding regions and gene promoters, but instead H-NS binds to particular DNA structures with the bridging between two DNA regions being important for both H-NS-mediated repression of gene expression and the compaction of the DNA. The structure of the DNA itself is affected by changes in the cellular environment and also by interactions with other proteins such as Fis, HU and sequence specific transcription factors. These proteins can antagonise or enhance the binding of H-NS with the DNA by moving regions of DNA where H-NS can bind either apart or bringing them closer together. The bridging of distant regions of DNA by H-NS is made possible by the self association of H-NS to form homodimers and larger oligomeric species although the size and importance of the larger oligomeric species remains the subject of some debate within the literature. However, there is increasing evidence that gene repression by H-NS is modulated by changes to both the H-NS DNA binding interaction itself and the self-association interactions of H-NS; both environmental changes and the formation of heteromeric complexes with H-NS-like proteins or proteins that mimic the H-NS oligomerisation domain such as the Hha/YmoA family of proteins affect the binding affinity of H-NS for DNA.

The biophysical characterisation of the oligomerisation domain and the nucleic acid binding domain of H-NS have been studied in this thesis to try to obtain a better understanding of the residues involved in the self-association interactions and the DNA binding interaction. The strategy taken was to separately examine the oligomerisation

domain and the nucleic acid binding domains so that amino acid substitutions and alteration of the buffer conditions could be directly attributed to a change in function of the isolated domain. The self-association of several C-terminally truncated H-NS constructs have been assessed by both size exclusion chromatography and analytical ultracentrifugation in order to determine the residues required for the formation of tetramers and higher order oligomers and therefore, establish the full extent of the H-NS oligomerisation domain.

Having established the residues that comprise the N-terminal oligomerisation domain of H-NS the effect of changing salt concentration on the oligomerisation interaction of H-NS was assessed using further analytical ultracentrifugation experiments and the effect of changing temperature on the coiled-coil dimerisation interaction was assessed using circular dichroism. Single point mutations within the oligomerisation domain have also been used to identify amino acids that are involved in the formation of tetramers and higher order oligomeric species.

The nucleic acid binding domains from *S. typhimurium* H-NS and *E. coli* StpA have been used to assess the different DNA binding properties of these isolated monomeric domains. NMR spectroscopy has been used to establish the amino acids from the H-NS and StpA nucleic acid binding domains that are responsible for DNA binding and to determine the affinity of the two domains for an AT-rich DNA molecule.

2 MATERIALS AND METHODS

The chapter describes the experimental methods used in this research. All chemicals were purchased from Sigma-aldrich or VWR unless otherwise stated.

2.1 PREPARATION OF DNA CONSTRUCTS

2.1.1 Purification of Plasmid DNA

Plasmid DNA was prepared from freshly transformed *E. coli* DH5α cultures grown overnight in 5 mL LB media containing antibiotic at 37°C and 210 cycles min^{-1} shaking. All plasmids were purified using the QIAprep® Spin Miniprep Kit (Qiagen) following the manufacturers instructions. Plasmid DNA was eluted from the spin column in sterile deionised water.

2.1.2 Cloning H-NS_{1-83} C21S into pET30aTEV Plasmid Vector

DNA encoding the H-NS gene from *Salmonella typhimurium* and theN-terminal 90 amino acids of this H-NS gene (encoding the oligomerisation domain and the flexible linker), both with a point mutation to replace cysteine 21 with serine, had been previously cloned into a pET14b vector (Novagen) between the NdeI and BamHI restriction sites and was kindly provided by Dr. C. Smyth. DNA encoding the StpA gene from *Escherichia coli* was cloned into pET14b using the same restriction sites and kindly provided by Dr. A. Petrovic.

pET30aTEV plasmid vector was provided by Dr. S. Djordjevic. This is a pET30a plasmid (Novagen), which has had the enterokinase recognition site replaced by a TEV protease recognition site.

When preparing protein samples for X-ray crystallography it is important to prepare as homogeneous a protein sample as possible. TEV protease is a more stringent protease than thrombin, cutting the polypeptide only at the TEV protease recognition site, so it was desirable to transfer the constructs for the oligomerisation domain into a vector that has a

TEV protease recognition site for cleavage of the His_6-tag to eliminate a potential source of heterogeneity. A further advantage to using TEV protease is that a recombinant TEV protease with a non-cleavable His_6-tag had been previously cloned into a pRET3a protein expression vector and could be expressed and purified when required, reducing the cost of the protein preparation and providing a simple means of removing the TEV protease after the cleavage of the His_6-tag is complete.

2.1.3 Calculating the Melting Temperature (T_M) of an Oligonucleotide Primer.

The T_M characterises the stability of the DNA hybrid formed between an oligonucleotide and its complementary strand. At T_M 50% of a given oligonucleotide is hybridized to its complementary strand. The T_M of the oligonucleotide primers was determined using MWG T_M calculation methods.

For oligonulceotides with less than 15 bases

$$T_M\left(^oC\right) = 2(n_A + n_T) + 4(n_G + n_C)$$

where n is the number of nucleotide bases of either A, T, G or C.

For oligonucleotides greater than 15 bases in length

$$T_M\left(^oC\right) = 69.3 + 41\left(\frac{(n_G + n_C)}{s}\right) - \left(\frac{650}{s}\right)$$

where s is the total number of bases in the sequence.

When calculating the T_M for oligonucleotides designed for site directed mutagenesis an alternative equation described in the QuikChange mutagenesis instructions (Stratagene) was used.

$$T_M\left(^oC\right) = 81.5 + 41\left(\frac{(n_G + n_C)}{s}\right) - \left(\frac{675}{s}\right) - 100\left(\frac{n_{mut}}{s}\right)$$

n_{mut} is the number of bases that are to be mutated.

2.1.4 Amplification of H-NS_{1-83} C21S DNA by Polymerase Chain Reaction

50 µL polymerase chain reactions containing 1x Thermopol reaction buffer (20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1 % Triton X-100 pH 8.8) (NEB), 200 µM dATP, 200 µM dTTP, 200 µM dGTP, 200 µM dCTP, 2 µM forward primer, 2 µM reverse primer, 10 ng template plasmid DNA and 2 enzyme units Vent DNA polymerase (NEB) were prepared. The forward (5'-CTA GTG CTA TGT CAT GAG CGA AGC ACT TAA AAT TCT G-3') and reverse primers (5'- CTG ATC GAA TTC TAC TAT TTA GCG GCA GC-3') used have calculated T_m values of 68.4°C and 65.3°C respectively. The nucleotides coloured blue indicate the position of the BspHI (forward primer) and EcoRI (reverse primer) restriction sites. For the cloning of H-NS_{1-83} C21S a pET14b plasmid with the DNA encoding *S. typhimurium* H-NS_{1-90} C21S was used as the template.

The polymerase chain reactions were performed in a Thermohybaid thermal cycler set to heat to 95°C for 5 min before 30 cycles of 95°C for 30 seconds, 63°C for 30 seconds and 72°C for 2 minutes. Polymerase chain reaction products were analysed by agarose gel electrophoresis and the appropriate sized DNA band excised and purified using the Eppendorf Perfectprep® Gel Cleanup kit.

2.1.5 Agarose Gel Electrophoresis

Agarose gel electrophoresis was used to separate and identify digested DNA fragments and PCR products. DNA fragments smaller than 1 kilobase (kb) were resolved on 1.5 % (w/v) agarose gels. Larger DNA fragments were separated using 1% (w/v) agarose gels. Agarose was dissolved in TBE buffer (45 mM Tris-borate, 1 mM EDTA) by heating and after cooling ethidium bromide was added to a final concentration of 0.5 µg mL^{-1}. Gels were cast and set in the apparatus provided. DNA samples were prepared by addition of 6X DNA loading buffer consisting of 0.1% (w/v) bromophenol blue, 0.1% (w/v) xylene cyanol and 30% (v/v) glycerol.

2.1.6 Restriction Digests

All restriction enzymes were supplied by NEB. The purified DNA product from the polymerase chain reaction was cut initially by 10 enzyme units of restriction endonuclease

BspHI in 1x NEB buffer 4 (20 mM Tris-acetate, 50 mM potassium acetate, 10 mM Magnesium Acetate pH 7.9) incubated overnight at 37°C. 20 enzyme units of EcoRI was then added and the reaction mixture incubated for a further 2 hours.

The pET30aTEV plasmid vector was double digested with NcoI and EcoRI restriction endonucleases in 1x EcoRI buffer (50 mM NaCl, 100 mM Tris-HCl, 10 mM $MgCl_2$, 0.025 % Triton X-100, pH 7.5) (NEB) for 2 hours at 37°C. Single restriction endonuclease enzyme reactions (BspHI or EcoRI only) were prepared as control reactions under the same conditions. The double digested polymerase chain reaction product or pET30aTEV was purified by agarose gel electrophoresis and the DNA band recovered from the gel using the Eppendorf Perfectprep® Gel Cleanup kit.

2.1.7 Ligation of Insert DNA into pET30aTEV Vector

A 20 µL ligation reaction was prepared in 1x T4 DNA ligase buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM Dithiothreitol, 25 µg ml^{-1} BSA, pH 7.5) with 100 ng NcoI and EcoRI digested pET30aTEV and approximately 20x molar excess of digested PCR product DNA. The reaction mixture was incubated at room temperature for 2 hours. A control reaction to test for the presence of undigested or religated plasmid template DNA, where the digested PCR product (insert DNA) was omitted from the reaction was also prepared using the same conditions.

2.1.8 Transformation of Ligation Products and Colony Testing

5 µL of the ligation reaction or the no insert ligation control reaction were transformed into competent *E. coli* DH5α cells as described in section 2.1.12. *E. coli* colonies were then tested for the presence of H-NS_{1-83} C21S DNA by polymerase chain reaction. Five colonies were selected and a sample of each colony dipped into 5 mL sterile luria bertani media supplemented with 100 µg mL^{-1} kanamycin. A 50 µL PCR reaction mixture was prepared as described in section 2.1.4 but the plasmid template DNA was replaced by deionised H_2O. This 10 µL solution was then pipetted into five thin walled PCR tubes. For each colony tested a sample of the colony was added to the PCR reaction mixture. The thermal cycling was performed as described previously in section 2.1.4 and the polymerase chain reaction products analysed by agarose gel electrophoresis. Plasmid

DNA was purified from the 5 mL overnight culture corresponding to colonies where a PCR product for insert DNA was observed and DNA sequencing was used to confirm the presence of the desired protein construct.

2.1.9 Mutagenesis of H-NS Plasmid Constructs

A modified version of the QuikChange™ Site Directed Mutagenesis Kit (Stratagene) protocol was used to make site specific point mutations in H-NS.

For the H-NS_{FL} C21S point mutants the pET14b plasmid with DNA encoding H-NS_{FL} C21S inserted between the NdeI and BamHI restriction endonuclease sites was used as the template plasmid. To prepare point mutants in the H-NS_{1-83} C21S construct, either to introduce stop codons that would truncate the C-terminal end of this protein or to introduce point mutations that replace charged amino acids with alanine, the pET30aTEV H-NS_{1-83} C21S plasmid (prepared in sections 2.1.2 to 2.1.7) was used as the template.

Primers containing the desired mutation or insertion were synthesised (MWG). The primers used in the quikchange reactions are described in Table 2-1, Table 2-2 and Table 2-3. PCR amplifications were performed in 50 µL reaction volumes containing 20 mM Tris HCl pH 8.8, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 0.2 mM $MgSO_4$, 0.1%Triton X-100, 0.1 mg mL^{-1} BSA, 200 nM forward primer, 200 nM reverse primers, 200 µM dNTP mix, 2.5 enzyme units Pfu Turbo DNA polymerase (Stratagene) and 10 ng template plasmid DNA. PCR amplification was performed using a Thermo Hybaid thermal cycler set to hold for 5 minutes at 95°C before performing 20 cycles of 95°C for 1 minute, 55°C for 1 minute and 68°C for 16 minutes. 10 enzyme units of DpnI restriction endonuclease (NEB) was added to each completed Quikchange reaction and incubated for 1 hour at 37°C to digest methylated template DNA. 10 µL samples of the reactions before and after incubation with DpnI were run on a 1.0% (w/v) agarose gel at 40 V for 1 hour to check for PCR products. 5 µL of DpnI treated PCR reaction product was then used to transform 50 µL *E. coli* DH5α competent cells (section 2.1.12). Plasmid DNA was purified from single colonies and DNA sequencing used to confirm the mutations.

Table 2-1. Mutagenesis primers for the preparation of different length oligomerisation domain constructs. Forward (_for) and reverse (_rev) primers used to introduce stop codons into the pET30aTEV H-NS_{1-83} C21S plasmid construct are shown. The nucleotide bases encoding the desired mutation are shown in red.

Mutation	Primer	Primer Sequence	Tm/°C
H-NS_{1-68} C21S	H-NS68_for	5'-CGT GAA ATG TTA ATT GCC GAC TAG ATT GAC CCG AAT GAA CTG CTG-3'	78.1
	H-NS68_rev	5'-CAG CAG TTC ATT CGG GTC AAT CTA GTC GGC AAT TAA CAT TTC ACG-3'	78.1
H-NS_{1-71} C21S	H-NS71_for	5'-ATT GCC GAC GGC ATT GAC TAG AAT GAA CTG CTG AAT AGC AT -3'	79.6
	H-NS71_rev	5'-ATG CTA TTC AGC AGT TCA TTC TAG TCA ATG CCG TCG GCA AT -3'	79.6
H-NS_{1-74} C21S	H-NS74_for	5'-GGC ATT GAC CCG AAT GAA TAG CTG AAT AGC ATG CGT GCC -3'	80.9
	H-NS74_rev	5'-GGC AGC CAT GCT ATT CAG CTA TTC ATT CGG GTC AAT GCC -3'	80.9
H-NS_{1-77} C21S	H-NS77_for	5'-CGA ATG AAC TGC TGA ATT AGA TGG CTG CCG CTA AAT AG -3'	74.2
	H-NS77_rev	5'-CTA TTT AGC GGC AGC CAT CTA ATT CAG CAG TTC ATT CG -3'	74.2
H-NS_{1-80} C21S	H-NS80_for	5'-GAA CTG CTG AAT AGC ATG GCT TAG GCT AAA TAG TAG AAT TCG AG -3'	76.1
	H-NS80_rev	5'-CTC GAA TTC TAC TAT TTA GCC TAA GCC ATG CTA TTC AGC AGT TC -3'	76.1

Table 2-2. Mutagenesis primers for the preparation charge mutants at the N-terminus of H-NS_{FL} C21S. Forward (_for) and reverse (_rev) primers used to introduce glutamate codons into the pET14b H-NS_{FL} C21S plasmid construct are shown. The nucleotide bases encoding the desired mutation are shown in red.

Protein Construct	Primer	Primer Sequence	Tm/°C
H-NS_{FL} R12E C21S	R12E_for	5'-GCA CTT AAA ATT CTG AAC AAC ATC GAA ACT CTT CGT GCG CAG GCA AG-3'	79.1
	R12E_rev	5'-CTT GCC TGC GCA CGA AGA GTT TCG ATG TTG TTC AGA ATT TTA AGT GC-3'	79.1
H-NS_{FL} R15E C21S	R15E_for	5'-CTG AAC AAC ATC CGT ACT CTT GAA GCG CAG GCA AGA GAA AGC-3'	78.8
	R15E_rev	5'-GCT TTC TCT TGC CTG CGC TTC AAG AGT ACG GAT GTT GTT CAG-3'	78.8

Table 2-3. Mutagenesis primers for the site directed mutagenesis replacing charged amino acids with alanine. Forward (_for) and reverse (_rev) primers used to introduce point mutations into H-NS plasmid constructs. The nucleotide bases encoding the desired mutation are shown in red.

Protein Construct	Primer	Primer Sequence	Tm/°C
H-NS R12A	R12A_for	5'-TAA AAT TCT GAA CAA CAT CGC TAC TCT TCG TGC GCA GGC-3'	75.4
	R12A_rev	5'-GCC TGC GCA CGA AGA GTA GCG ATG TTG TTC AGA ATT TTA-3'	75.4
H-NS R15A	R15A_for	5'-CTG AAC AAC ATC CGT ACT CTT GCT GCG CAG GCA AG-3'	75.9
	R15A_rev	5'-CTT GCC TGC GCA GCA AGA GTA CGG ATG TTG TTC AG-3'	75.9
H-NS E74A	E74A_for	5'-GCA TTG ACC CGA ATG CAC TGC TGA ATA GCA TG-3'	77.8
	E74A_rev	5'-CAT GCT ATT CAG CAG TGC ATT CGG GTC AAT GC-3'	77.8

2.1.10 Growth of *E. coli* Cultures

E. coli cells were grown in Luria Bertani (LB) media (10 g L^{-1} Bacto tryptone, 5 g L^{-1} yeast extract, 10 g L^{-1} NaCl) supplemented with the appropriate antibiotic(s): 50 µg mL^{-1} chloramphenicol for pLysS plasmid, 100 µg mL^{-1} carbenicillin for pET14b plasmid constructs (H-NS_{FL} C21S, H-NS_{90-137}, H-NS_{FL} R12E/C21S, H-NS_{FL} R15E/C21S, $StpA_{FL}$, $StpA_{91-134}$) or 50 µg mL^{-1} kanamycin for pET30aTEV plasmid constructs (H-NS_{1-68} C21S, H-NS_{1-71} C21S, H-NS_{1-74} C21S, H-NS_{1-77} C21S, H-NS_{1-80} C21S, H-NS_{1-83} C21S, H-NS_{1-83} R12A/C21S and H-NS_{1-83} C21S/E74A)

The LB media was supplemented with 15 g L^{-1} agar and the appropriate antibiotic(s) for the preparation of agar plates.

2.1.11 Preparation of Competent *E. coli* Cells

The following procedure was used to prepare competent cells of *E. coli* DH5α, *E. coli* BL21 Star (DE3) pLysS and *E. coli* B834 pLysS cells. When using cells containing the pLysS plasmid the LB media was supplemented with 50 µg mL^{-1} chloramphenicol. *E. coli* DH5α cells were grown in LB without any antibiotics present.

A single colony of *E. coli* cells was used to inoculate 5 mL LB prior to overnight incubation at 37°C, 210 cycles min^{-1} shaking. The overnight culture was diluted 10 fold in LB and incubated as before, until an optical density at 600 nm of 0.7 absorbance units was reached. The cells were collected by centrifugation at 4000 rpm, 4°C for 20 minutes in a PK130R centrifuge (ALC) fitted with a T535 rotor. The cell pellet was resuspended in 17 mL sterile transformation buffer I (30 mM potassium acetate pH 5.8, 100 mM RbCl, 50 mM $MnCl_2$, 10 mM $CaCl_2$, 15% (w/v) glycerol). After incubating the suspension on ice for 2 hours the cells were spun down by centrifugation as described previously. The cell pellets were resuspended in 4 mL sterile transformation buffer II (10 mM Mops pH 6.8, 10 mM RbCl, 75 mM $CaCl_2$, 15% (w/v) glycerol). 100 µL aliquots were stored at -80°C until required.

2.1.12 Transformation of *E. coli* Cells

50 µL of competent *E. coli* cells incubated on ice for 30 minutes with either 5 µL of Quikchange reaction product (nicked plasmid DNA), 5 µL of ligation reaction product or 0.5 µL of supercoiled plasmid DNA. The cells were heat shocked for 45 seconds at 42°C. 450 µL of NZY+ media (10 g L^{-1} NZ amine, 5 g L^{-1} yeast extract, 5 g L^{-1} NaCl, 20% (w/v) glucose, 12.5 mM $MgCl_2$, 12.5 mM $MgSO_4$) was then added before incubating the cell suspension at 37°C for 1 hour. 250 µL of the cell suspension was then spread onto a LB agar plate supplemented with the appropriate antibiotic(s) and the plate incubated for 16 hours at 37°C. Single colonies were observed after this incubation period.

2.1.13 Preparation of Glycerol Stocks of Transformed *E. coli* Cells

5 mL of sterile luria bertani media supplemented with the appropriate antibiotic(s) was inoculated with a single colony of the *E. coli* cells transformed with the desired plasmid DNA. This culture was grown at 37°C, 210 cycles min^{-1} shaking for 16 hours. 850 µL of this cell culture was then added to 150 µL sterile glycerol in a sterile microfuge tube. After mixing the solution by pipette the tubes were stored at -80°C until required.

2.2 PROTEIN EXPRESSION

2.2.1 Expression of Unlabelled H-NS and StpA Protein Constructs

A single colony of *E. coli* BL21 Star (DE3) pLysS, transformed with a pET vector plasmid DNA encoding the desired protein (pET14b for H-NS_{FL} C21S, H-NS_{FL} R12E/C21S, H-NS_{FL} R15E/C21S, H-NS_{90-137}, $StpA_{FL}$ or $StpA_{91-134}$; pET30aTEV for H-NS_{1-68} C21S, H-NS_{1-71} C21S, H-NS_{1-74} C21S, H-NS_{1-77} C21S, H-NS_{1-80} C21S, H-NS_{1-83} C21S, H-NS_{1-83} R12A/C21S and H-NS_{1-83} C21S/E74A) was used to inoculate 100 mL of sterile LB supplemented with 50 µg mL^{-1} chloramphenicol and either 100 µg mL^{-1} carbenicillin (pET14b plasmids) or 50 µg mL^{-1} kanamycin (pET30aTEV plasmids) in a 250 mL conical flask. The flask was incubated at 37°C, 210 cycles min^{-1} shaking for 16 hours. 10 mL of this culture was used to inoculate 2 L baffled flasks containing 500 mL sterile LB supplemented with the appropriate antibiotics.

The cultures were incubated until an optical density at 600nm of 1.0 absorbance unit is reached. At this point 1 M sterile IPTG was added to give a final concentration of 1 mM and the culture incubated for a further 4 hours. The cells were collected by centrifugation at 5000 rpm, 4°C for 15 minutes using a Sorvall SLA-3000 rotor and the supernatent discarded. The cells were stored at -20°C until required.

2.2.2 Protein Expression for ^{15}N-Labelled Protein Constructs

It was necessary to use controlled growth media for the uniform isotopic enrichment of expressed protein with ^{15}Nitrogen (^{15}N) for the preparation of NMR protein samples. For this purpose a single colony of *E. coli* BL21 Star (DE3) pLysS transformed with the plasmid for the protein of interest was used to inoculate 5 mL of LB. After incubating this culture at 37°C, 210 cycles min^{-1} shaking for 6 hours the cells were collected by centrifugation at 4000 rpm, 4°C for 20 minutes. The supernatent was discarded and the cells resuspended in 100 mL of sterile ^{15}N media.

To prepare ^{15}N media, the following stock solutions were prepared.

- M9 minimal media: 1 L solution of 42.3 mM Na_2HPO_4, 14.7 mM KH_2PO_4, 8.56 mM NaCl and 18.7 mM $(^{15}NH_4)_2SO_4$ pH 7.2 was prepared and autoclaved to sterilise.
- 1 M $MgCl_2$. Sterile filtered through a 0.2 µm filter.
- 100 mM $CaCl_2$. Sterile filtered through a 0.2 µm filter.
- 20% (w/v) D-glucose. Sterile filtered through a 0.2 µm filter.
- 100 mg mL^{-1} carbenicillin. Sterile filtered through a 0.2 µm filter.
- 50 mg mL^{-1} chloramphenicol in ethanol.
- 5000x micronutrient solution: 50 mM $FeCl_2$, 2 mM $CuCl_2$, 2 mM Sodium molybdate, 2 mM $NiCl_2$, 10 mM $MnCl_2$, 10 mM $ZnCl_2$, 20 mM $CaCl_2$, 2 mM $CoCl_2$, 2 mM H_3BO_3. Sterile filtered through a 0.2 µm filter.
- 10000x vitamin solution: 28.7 µM choline chloride, 11.3 µM folic acid, 10.5 µM pantothenic acid, 40.9 µM nicotinamide, 55.5 µM myo-inositol, 24.6 µM pyridoxal HCl, 14.8 µM thiamine HCl, 1.33 µM riboflavin and 40.9 µM biotin. Sterile filtered through a 0.2 µm filter.

The supplemented minimal media was then prepared by adding 2 mL 1 M $MgCl_2$, 100 µL 100 mM $CaCl_2$, 10 mL 20% glucose, 1 mL 100 mg mL^{-1} carbenicillin, 1 mL 50 mg mL^{-1} chloramphenicol, 200 µL 5000x micronutrient solution and 100 µL 10000x vitamin solution to a litre of M9 minimal media.

A 100 mL *E. coli* culture was then incubated at 37°C, 210 cycles min^{-1} shaking overnight. The following day 10 mL of this overnight culture was used to inoculate 2L baffled flasks containing 500 mL of sterile ^{15}N media. These cultures were grown at 37°C, 210 cycles min^{-1} shaking until the optical density at 600 nm reached 0.7 absorbance units. At this point 500 µL 1 M IPTG was added to each flask to induce expression of the protein of interest. The cultures were incubated for a further 4 hours before the cells were collected by centrifugation at 5000 rpm, 4°C for 15 minutes in a Sorvall SLA-3000 rotor. The supernatent was discarded and the cells stored at -20°C.

2.2.3 Protein Expression of Selenomethionine Labelled Protein Constructs

In the absence of a suitably homologous model protein to use for molecular replacement interpretation of an X-ray diffraction pattern it is necessary to prepare a protein sample under conditions where the methionine residues are replaced by selenomethionine. The selenium allows you to perform single anomalous dispersion (SAD) or multiple anomalous dispersion (MAD) experiments. To prepare a selenomethionine protein sample a methionine auxotroph *E. coli* B834 (DE3) *pLysS* strain needs to be used to express the protein under growth conditions where the only source of methionine is L-selenomethionine.

The growth conditions for the preparation of selenomethionine labelled protein were based on a method developed by Neidhardt F. et al. (Neidhardt *et al.*, 1974). The following stock solutions were prepared to make media suitable for the expression of selenomethionine labelled protein.

- 10x MOPS concentrate: 400 mM Mops pH 7.4, 40 mM Tricine, 0.1 mM $FeSO_4$, 95 mM NH_4Cl, 2,76 mM K_2SO_4, 5µM $CaCl_2$, 5.3 mM $MgCl_2$, 500 mM NaCl. Autoclaved to sterilise.
- 20% D-glucose. Sterile filtered through a 0.2 µm filter.

- 5x Amino acid stock solution: 4 mM L-alanine, 2.5 mM L-arginine, 2 mM L-asparagine, 2 mM L-aspartate, 0.5 mM L-cysteine, 3 mM L-glutamate, 3 mM L-glutamine, 4 mM glycine, 1 mM L-histidine, 4 mM L-isoleucine, 8 mM L-leucine, 4 mM L-lysine, 2 mM L-phenylalanine, 2 mM L-proline, 18.3 mM L-serine, 2 mM L-threonine, 0.5 mM L-tryptophan, 2 mM L-tyrosine and 6 mM L-valine. Sterile filtered through a 0.2 µm filter.
- 132 mM K_2HPO_4 sterile filtered through a 0.2 µm filter.
- 100 mg mL^{-1} carbenicillin sterile filtered through a 0.2 µm filter.
- 50 mg mL^{-1} chloramphenicol dissolved in ethanol.
- 5000x Micronutrient stock solution: 50 mM $FeCl_2$, 2 mM $CuCl_2$, 2 mM Sodium molybdate, 2 mM $NiCl_2$, 10 mM $MnCl_2$, 10 mM $ZnCl_2$, 20 mM $CaCl_2$, 2 mM $CoCl_2$, 2 mM H_3BO_3.
- 2000x Vitamin stock solution: 2 mM thaimine, 1 mM pantothenic acid, 2.3 mM p-hydroxybenzoic acid, 2 mM p-aminobenzoic acid, 1.75 mM 2,3-dihydroxybenzoic acid.

For half a litre of seleno-methionine media 50 mL of 10x MOPS concentrate was added to 330 mL sterile autoclaved dH_2O in a sterile baffled 2L flask. To this solution was added 15 mL 20% (w/v) sterile filtered glucose, 100 mL 5x amino acid stock solution, 5 mL 132 mM K_2HPO_4, 500 µL 100 mg mL^{-1} carbenicillin, 500 µL 50 mg mL^{-1} chloramphenicol 250 µL 2000x vitamin stock solution, 100 µL 5000x micronutrient stock solution and 25mg L-selenomethionine. Two litres of selenomethionine media were prepared in total for the preparation of a selenomethionine labelled H-NS_{1-83} C21S protein sample.

A single colony of *E. coli* B834 pLysS transformed with the pET14b-H-NS_{1-83} C21S plasmid was used to inoculate 5 mL of LB. After incubating this culture at 37°C, 210 cycles min^{-1} shaking the cells were collected by centrifugation at 4000 rpm, 4°C for 20 minutes. These cells were then resuspended in 100 mL sterile selenomethionine media and incubated overnight. The following day 20 mL of this culture was used to inoculate 2 L baffled flasks containing 500 mL sterile selenomethionine media prepared immediately prior to use. These culture was grown until an optical density of 1.0 absorbance units was reached before 500 µL 1 M IPTG was added to each flask to induce expression of the H-NS_{1-83} C21S protein. The cultures were incubated for a further 4 hours before the cells were collected by centrifugation at 5000 rpm, 4°C for 15 minutes in a Sorvall SLA-3000 rotor. The supernatent was discarded and the cells stored at -20°C.

2.3 PROTEIN PURIFICATION

All proteins were cloned into protein expression vectors to enable production of a polypeptide with a protease cleaveable N-terminal His_6-tag upon induction with IPTG. Proteins cloned into pET30aTEV vector had a TEV protease cleavable His_6-tag whereas the proteins cloned into the pET14b vector had a thrombin cleavable His_6-tag. All constructs that include the N-terminal oligomerisation domain of H-NS have been expressed with an N-terminal His_6-tag to aid purification, which prevents the post-translational cleavage of the N-terminal methionine by an exoprotease.

2.3.1 Cell Lysis

The collected *E. coli* cells were resuspended on ice in 20 mM potassium phosphate pH 7.0, 500 mM NaCl, 1 mM NaN_3, 1 mM AEBSF, 1 mM Benzamidine, 10μM E64. The cell suspension was then sonicated for eleven cycles of 10 seconds pulses separated with 20 second intervals, on ice to lyse the cells. The insoluble components were removed by centrifugation at 18000 rpm, 4°C for 30 minutes in a Sorvall SS34 rotor.

2.3.2 Talon Metal Affinity Resin Purification of His_6-tagged Proteins

The supernatent, containing the soluble protein components from the cell lysate, was then loaded at 2 mL min^{-1} onto a 5 mL Talon metal affinity resin column (BD Biosciences) pre-equilibrated with 20 mM potassium phosphate, pH 7.0, 500 mM NaCl, 1 mM NaN_3 using an AKTAprime FPLC system (GE Healthcare). A UV detector measuring absorbance at 280 nm was used to monitor the protein eluting from the column. Once the UV signal had returned to the baseline the column was washed with 200 mL 20 mM potassium phosphate, pH 7.0, 500 mM NaCl, 10 mM imidazole, 1 mM NaN_3 to remove weakly bound contaminating proteins. The His_6-tagged protein was eluted using 20 mM potassium phosphate, pH 7.0, 500 mM NaCl, 150 mM imidazole, 1 mM NaN_3. The protein concentration was estimated using the Bradford assay method. The imidazole and any protein impurities were removed from the Talon column by washing with 100 mL 6 M guanidine pH 5.0, 1% (v/v) Triton-X100 followed by 100 mL 20 mM Mes pH 5.0, 100 mM NaCl and 100 mL dH_2O.

2.3.3 Proteolytic Cleavage of His_6-tags

Either 2.5 enzyme units of high activity bovine thrombin (Calbiochem) per milligram of purified H-NS or StpA protein or 0.1 mg TEV protease per milligram of purified H-NS or StpA protein was added to the eluted protein solution and the protein solution sterile filtered using a 0.2 µm filter to remove any microbial contaminants. The protein solution was then dialysed against two litres of 20 mM potassium phosphate, pH 7.0, 500 mM NaCl, 1 mM NaN_3. The dialysis solution was changed for a fresh two litres twice during overnight incubation at 4°C. SDS PAGE analysis was used to confirm that the proteolytic cleavage reaction was complete.

2.3.4 Removal of His_6-tags, Uncleaved Protein and Protease Contaminant

After the proteolysis reaction is complete the protein solution was passed back through the talon column after pre-equilibration of the resin with 20 mM potassium phosphate, pH 7.0, 500 mM NaCl, 1 mM NaN_3, to remove any cleaved His_6-tags, uncleaved protein and TEV protease (TEV protease has an uncleavable N-terminal His_6-tag).

When thrombin was used for proteolytic cleavage of the His_6-tag 1 mL of pre-equilibrated Benzamidine-sepharose resin was incubated with the protein solution for 30 minutes at 4°C with gentle stiring. The suspension was then centrifuged to pellet the Benzamidine-sepharose with the thrombin bound and the supernatent containing the protein of interest was carefully removed. This protein solution was then passed through a 0.2 µm sterile filter to ensure all of the resin was removed.

2.3.5 Q-Sepharose Anion Exchange

This anion exchange purification procedure was applied to protein constructs of the N-terminal oligomerisation domain: H-NS_{1-68} C21S, H-NS_{1-71} C21S, H-NS_{1-74} C21S, H-NS_{1-77} C21S, H-NS_{1-80} C21S, H-NS_{1-83} C21S, H-NS_{1-83} R12A/C21S and H-NS_{1-83} C21S/E74A. After removal of the cleaved His_6-tags and protease contaminant the protein solution was diluted with 20 mM potassium phosphate, pH 7.0, 1 mM NaN_3 1 mM EDTA to reduce the concentration of NaCl in the buffer to 50 mM. This protein solution was then loaded onto a column packed with 20 mL Q-sepharose fast flow resin (GE Healthcare) pre-equilibrated

with 20 mM potassium phosphate, pH 7.0, 1 mM NaN_3, 1 mM EDTA. After washing the resin the bound protein was eluted using a linear NaCl gradient from no NaCl to 1 M NaCl. Fractions containing eluted protein were analysed by SDS PAGE and those containing the desired protein were pooled and stored at 4°C.

2.3.6 SP-Sepharose Cation Exchange

This cation exchange purification procedure was applied as a second purification step for the purification of the following protein constructs: H-NS_{FL} C21S, $StpA_{91-134}$, H-NS_{90-137}. After removal of the cleaved His_6-tags and protease contaminant the protein solution was diluted with 20 mM potassium phosphate, pH 7.0, 1 mM NaN_3, 1 mM EDTA to reduce the concentration of NaCl in the buffer to 50 mM. This protein solution was then loaded onto a column packed with 20 mL SP-sepharose fast flow resin (GE Healthcare) pre-equilibrated with 20 mM potassium phosphate, pH 7.0, 1 mM NaN_3, 1 mM EDTA. After washing the resin the bound protein was eluted using a linear NaCl gradient from 0 mM to 1 M NaCl. Fractions containing eluted protein were analysed by SDS PAGE and those containing the desired protein were pooled and stored at 4°C.

2.3.7 Preparative Size Exclusion Chromatography

Preparative Size Exclusion Chromatography was applied to all protein constructs. A column packed with 120 mL Superdex 75 resin (GE Healthcare) was pre-equilibrated with 20 mM potassium phosphate, pH 7.0, 300 mM NaCl. Using a vivaspin concentrator with appropriate MW cut off the protein solution was concentrated to less than 5 mL final volume before being loaded onto the Superdex column. The column was run at a flow rate of 1 mL min^{-1} and 2 mL fractions were collected. Protein elution was detected by UV absorbance at 280nm.

2.4 PRELIMINARY PROTEIN ANALYSIS

2.4.1 Protein Concentration Determination

The concentrations of all constructs used in this study were determined using a Cary 50 UV-visible spectrophotometer measuring the absorbance of the protein at 280nm in a 1 cm pathlength quartz cuvettes (Hellma). Molar extinction coefficients were calculated based on the amino acid composition of the protein (Gill and von Hippel, 1989) and are shown in Table 2-4. The protein concentration was calculated using the Beer-Lambert law

$$A_{280} = \varepsilon_{280}.c.l$$

where A_{280} is the absorbance of the protein solution at 280 nm, ε_{280} is the molar extinction coefficient of the protein at 280 nm, c is the molar protein concentration (M^{-1} cm^{-1}) and l is the pathlength of the sample in centimetres.

2.4.2 SDS-PAGE

Protein samples at each stage of the purification procedure were analysed by Tris-tricine sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) to assess the purity and stability of the protein that was being purified. The gel consists of a 4% polyacrylamide stacking layer and a 13.3% polyacrylamide resolving layer. The gels were prepared using the amounts of the stock solutions shown in Table 2-5. 3 mL of resolving gel solution was added to each gel mould before it set and 200µL isopropanol added to ensure that the gel depth was even across the gel. Once set the isopropanol was removed and the stacking gel solution was prepared. The gel mould was filled and a stacking comb added to create sample wells. Electrophoresis was carried out on the Mini-Protean III system (Biorad) linked to a PowerPac 300 power supply (Biorad) using 100 mM Tricine, 100 mM Tris Base, 0.1% SDS as the running buffer. The protein samples were mixed with 5x SDS sample buffer (0.5 M Tris-base pH 6.8, 10% (w/v) SDS, 0.5% (w/v) bromophenol blue, 50% (v/v) glycerol) and heated to 95°C for 5 minutes. The samples were then centrifuged at 13000 rpm for 2 minutes to remove any insoluble material before a sample of the solution was loaded into the gel. The samples were electrophoresed concurrently with Precision Plus protein molecular weight standards (Biorad).

Table 2-4. Molar extinction coefficients for the protein constructs used in this report.

Protein Construct	Molar Extinction Coefficient at 280nm (M^{-1} cm^{-1})	Molecular Mass Before His_6-tag Cleavage (g mol^{-1})	Molecular Mass After His_6-tag Cleavage (g mol^{-1})
H-NS_{FL} C21S	9530	17705.9	15823.9
StpAfl	12660	17510.8	15628.0
H-NS_{1-58} C21S	0	8907.0	7024.9
H-NS_{1-65} C21S	1280	9856.1	7974.0
H-NS_{1-68} C21S	1280	13240.8	8104.1
H-NS_{1-71} C21S	1280	13526.1	8389.4
H-NS_{1-74} C21S	1280	13866.4	8729.8
H-NS_{1-77} C21S	1280	14206.8	9070.2
H-NS_{1-80} C21S	1280	14496.2	9359.5
H-NS_{1-83} C21S	1280	14766.5	9629.9
H-NS_{90-137}	8480	7608.4	5726.4
$StpA_{1-65}$	5500	10120.4	8238.4
$StpA_{91-134}$	6790	7170.0	5287.9

Table 2-5. Protocol for the preparation of a 13.3% polyacrylamide gel for SDS-PAGE.

Stock Solution	13.3% Resolving Gel/µL	4% Stacking Gel/µL
3x Gel Buffer (3 M Tris pH 8.45, 0.3% SDS, 30% Glycerol	3000	2000
30% acrylamide solution (37.5:1 ratio of acrylamide:bisacrylamide	4000	800
10% ammonium persulphate	100	100
N,N,N',N'-Tetramethylethylenediamine (Temed)	10	10
Deionised Water	1890	3090

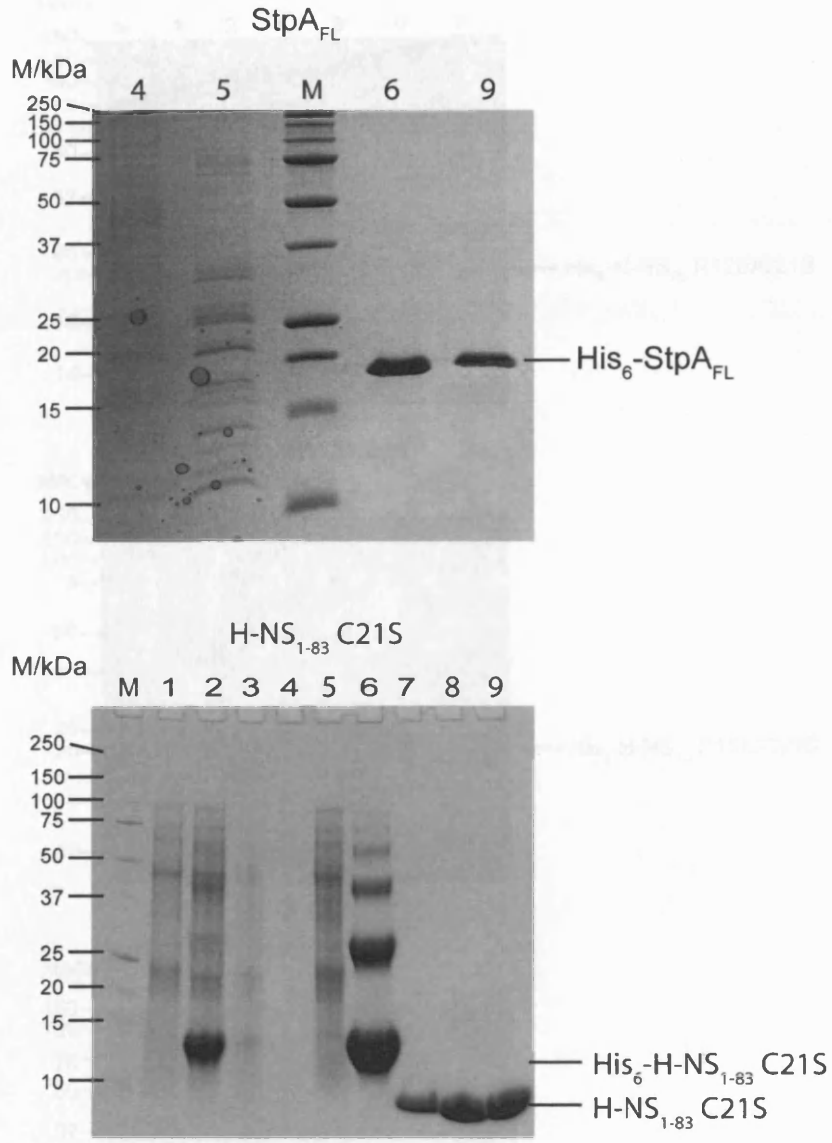


Figure 2-1. SDS PAGE analysis for the protein purification of $StpA_{FL}$, and H-NS_{1-83} C21S. The size of the molar mass markers in kDa is indicated to the left of each gel. The protein samples loaded are: M. Molar mass markers; 1. pre IPTG induction; 2. post IPTG induction; 3. insoluble pellet; 4. Talon column flow through; 5. 10 mM imidazole wash; 6. 150 mM imidazole elution; 7. TEV cleavage; 8. post Q-sepharose column; 9. post size exclusion chromatography column.

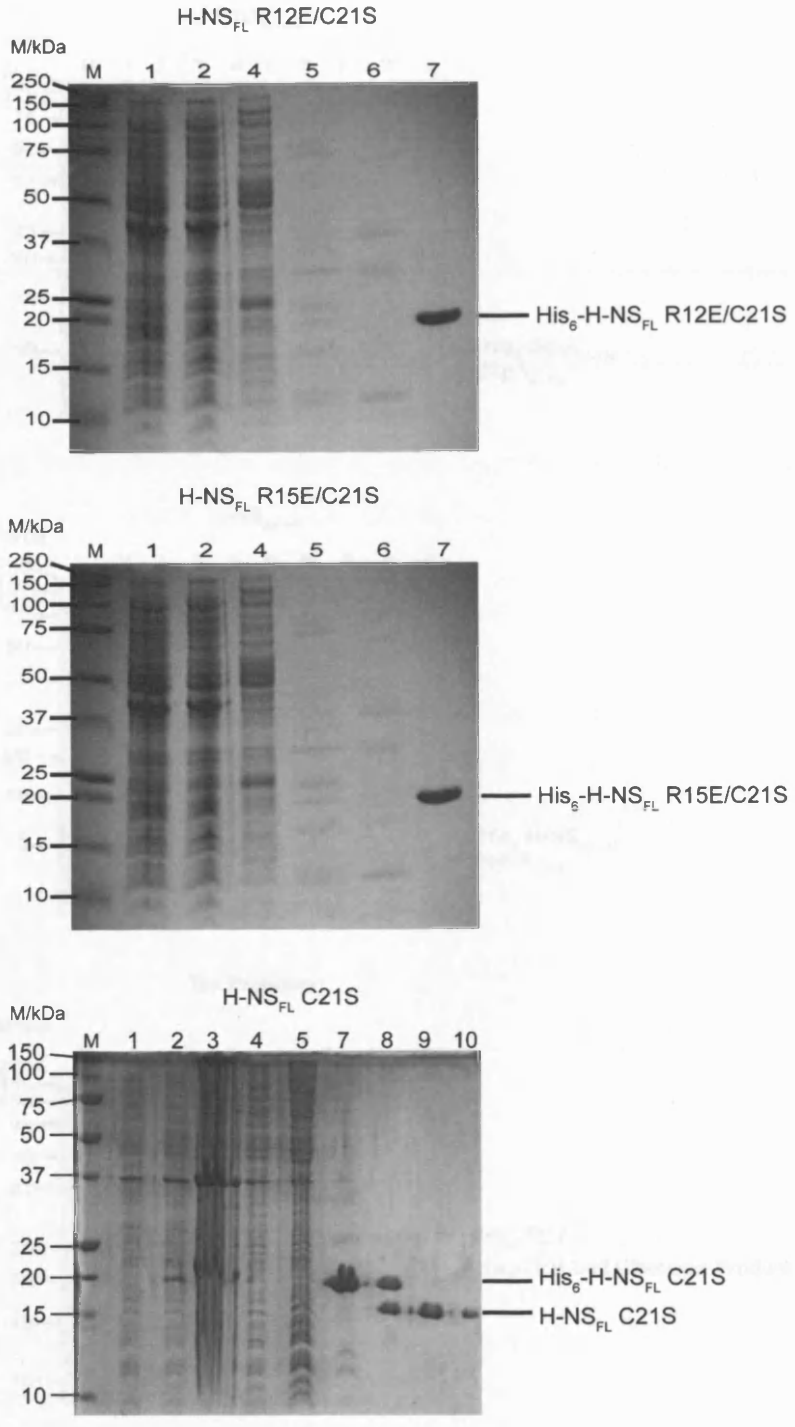


Figure 2-2. SDS PAGE analysis for the protein purification of H-NS$_{FL}$ 12E/C21S, H-NS$_{FL}$ 15E/C21S and H-NS$_{FL}$ C21S recombinant proteins. The size of molar mass markers is indicated to the left of each gel. The protein samples were as follows: M. Molar mass markers; 1. pre IPTG induction; 2. post IPTG induction; 3. insoluble pellet; 4. Talon column flow through; 5. 10 mM imidazole wash; 6. 40 mM imidazole wash; 7. 150 mM imidazole elution; 8. thrombin cleavage; 9. Post SP-sepharose column; 10. Post size exclusion chroamtography column

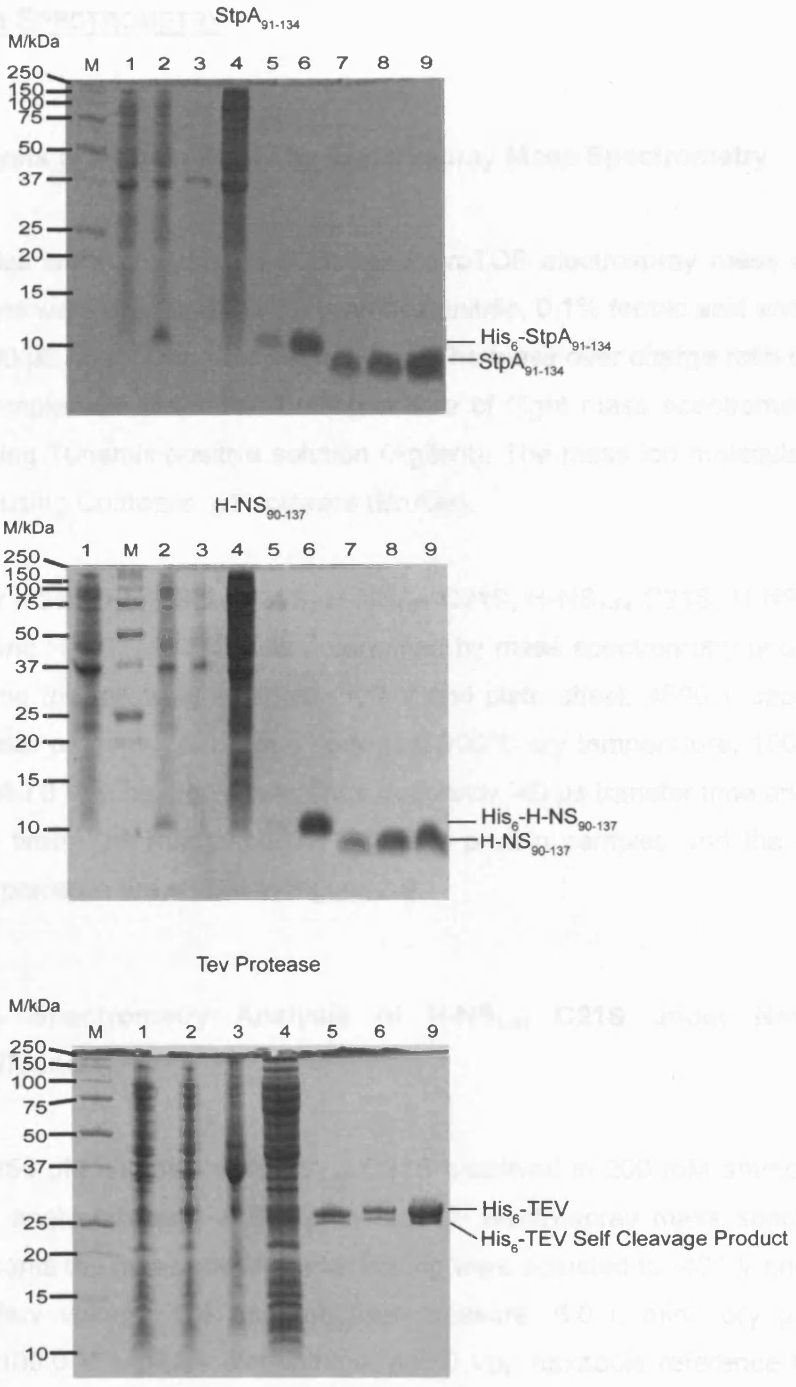


Figure 2-3. SDS PAGE analysis for the protein purification of $StpA_{91-134}$, $H\text{-}NS_{90-137}$ and TEV protease. The size of the molar mass markers in kDa is indicated to the left of each gel. The protein samples loaded are: M. Molar mass markers; 1. pre IPTG induction; 2. post IPTG induction; 3. insoluble pellet; 4. Talon column flow through; 5. 10 mM imidazole wash; 6. 150 mM imidazole elution; 7. thrombin cleavage; 8. post SP-sepharose column; 9. post size exclusion chromatography column.

2.5 MASS SPECTROMETRY

2.5.1 Analysis of Protein Purity by Electrospray Mass Spectrometry

Protein samples were analysed in a Bruker MicroTOF electrospray mass spectrometer. Protein samples were prepared in 50% (v/v) acetonitrile, 0.1% formic acid and injected at a flow rate of 200 µL $hour^{-1}$ using a syringe pump. The mass over charge ratio of the positive ions in the sample was determined using a time of flight mass spectrometer calibrated before use using Tunemix positive solution (Agilent). The mass ion molecular weight was deconvoluted using Compass 1.1 software (Bruker).

The molecular mass of H-NS_{1-68} C21S, H-NS_{1-71} C21S, H-NS_{1-74} C21S, H-NS_{1-77} C21S, H-NS_{1-80} C21S and H-NS_{1-83} C21S was determined by mass spectrometry under denaturing conditions using the following settings: -500 V end plate offset, 4500 V capillary voltage, 0.4 bar nebuliser pressure, 5.0 L min^{-1} dry gas, 200°C dry temperature, 150.0 V capillary exit voltage, 350.0 Vpp hexapole reference frequency, 40 µs transfer time and 18.5 µs pre pulse storage time. The mass spectra for these protein samples and the deconvoluted mass ion interpretation are shown in Figure 2-9.

2.5.2 Mass Spectrometry Analysis of H-NS_{1-83} C21S under Non-denaturing Conditions

500 µM and 250 µM samples of H-NS_{1-83} C21S dissolved in 200 mM ammonium acetate pH 7.0. were analysed using a Bruker MicroTOF electrospray mass spectrometer. For these experiments the mass spectrometer setting were adjusted to -400 V end plate offset, 4500 V capillary voltage, 0.4 bar nebuliser pressure, 6.0 L min^{-1} dry gas, 37°C dry temperature, 100.0 V capillary exit voltage, 400.0 Vpp hexapole reference frequency, 60 µs transfer time and 15 µs pre pulse storage time.

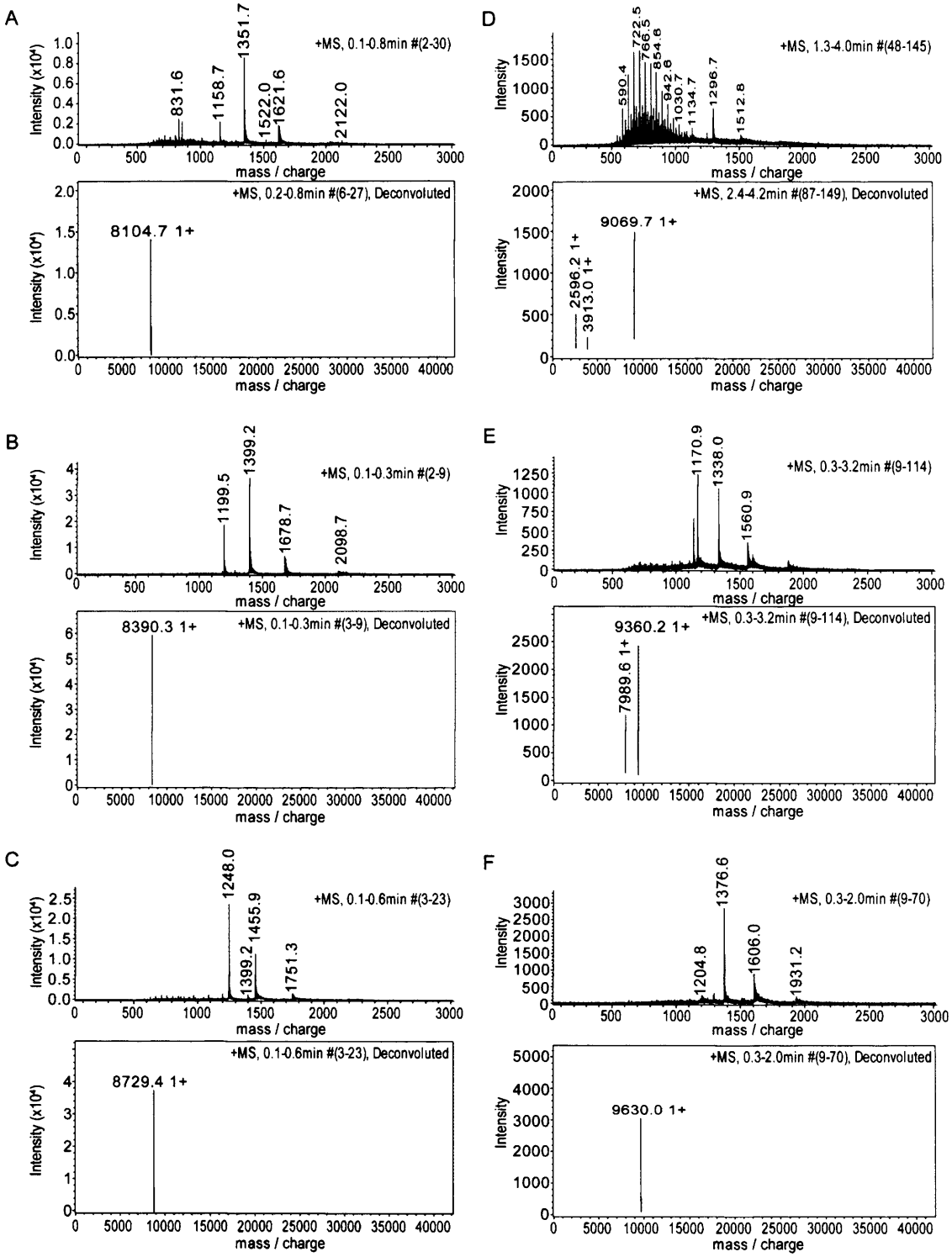


Figure 2-4. Mass spectrometry analysis of the oligomerisation domain constructs of H-NS. The diagrams show the charged species identified by the mass spectrometer (above) and the deconvoluted mass ion interpretation (below) for A. H-NS_{1-68} C21S, B. H-NS_{1-71} C21S, C. H-NS_{1-74} C21S, D. H-NS_{1-77} C21S, E H-NS_{1-80} C21S and F. H-NS_{1-83} C21S.

2.6 ANALYTICAL SIZE EXCLUSION CHROMATOGRAPHY

2.6.1 Analysis of H-NS Truncation and Point Mutation Constructs by Analytical Size Exclusion Chromatography.

Analytical size exclusion chromatography was used to assess the size and heterogeneity of the various protein constructs of the H-NS oligomerisation domain and full length protein. A pre-equilibrated superose 12HR column (G.E. life sciences) was connected to an $AKTA_{FPLC}$ system. 100 µL protein samples were loaded onto the column for each experiment. A range of concentrations for each protein were tested dictated by the solubility or amount of protein available from the purification. The concentrations tested are stated in the appropriate results section.

For the analysis of the N-terminal oligomerisation domain constructs ($H\text{-}NS_{1\text{-}68}$ C21S, $H\text{-}NS_{1\text{-}71}$ C21S, $H\text{-}NS_{1\text{-}74}$ C21S, $H\text{-}NS_{1\text{-}77}$ C21S, $H\text{-}NS_{1\text{-}80}$ C21S, $H\text{-}NS_{1\text{-}83}$ C21S and all of the $H\text{-}NS_{1\text{-}83}$ C21S point mutants, a buffer containing 20 mM potassium phosphate pH 7.0, 300 mM NaCl, 1 mM EDTA was used for the analysis.

Protein samples of purified $H\text{-}NS_{1\text{-}83}$ C21S were also tested in 20 mM potassium phosphate pH 7.0, 1 mM EDTA to assess the affect of low ionic strength on the oligomeric state of the protein.

For the experiments using $H\text{-}NS_{FL}$ C21S, $H\text{-}NS_{FL}$ R11E/C21S and $H\text{-}NS_{FL}$ R15E/C21S a buffer containing 20 mM potassium phosphate, 600 mM NaCl, 1 mM EDTA was used for the analysis. The $H\text{-}NS_{FL}$ R15E/C21S protein precipitates in 20 mM potassium phosphate, 300 mM NaCl, 1 mM EDTA conditions at concentrations above 100 µM so a higher NaCl concentration was required to make this construct more soluble.

2.6.2 Calibration of the Analytical Size Exclusion Chromatography Column

In order to assess the apparent molar mass of the various H-NS constructs the superose 12HR column was calibrated using seven globular protein size standards ranging in molar mass from 13.7 kDa to 660 kDa. The size standards were purchased from Sigma or G.E. life sciences. Each protein sample was prepared in the solvent used to equilibrate the

column supplemented with 10 mM DTT to prevent disulphide bond formation. The superose 12HR column was calibrated in both 20 mM potassium phosphate pH 7.0, 300 mM NaCl, 1 mM EDTA (Figure 2-5) and 20 mM potassium phosphate pH 7.0, 1 mM EDTA (Figure 2-6) buffer conditions. A plot of $\log_{10}$ molar mass against the observed elution volume was used to calibrate the column. Only the points that obey a linear relationship between log molar mass and the elution volume were used to determine the gradient and y intercept for the calibration curve.

2.7 CIRCULAR DICHROISM

Circular dichroism (CD) spectroscopy was carried out on protein samples to confirm the folded state of the protein using a 202SF AVIV CD spectropolarimeter. The CD spectropolarimeter calibration was checked by measuring the CD signal at 192.5 nm and 290 nm of a 1 mg mL^{-1} sample of (1S)-(+)-10-camphorsulphonic acid prepared in a 1 mm pathlength quartz cuvette. The absorbance profile of the protein solution and reference buffer solution respectively, for the wavelength range of 190 nm to 260 nm was initially tested using a 0.01 cm pathlength quartz cuvette using a Cary 50 UV-Visible spectrophotometer with deionised H_2O as the reference. After determining the absorbance properties of both the protein and buffer components in this wavelength range the optimal pathlength cuvette was then selected to give the most accurate CD measurement; the cuvette with a pathlength that would maximise the absorbance measured whilst keeping the signal below the 1.5 absorbance unit limit of the detector was chosen. The CD signal, θ, at 25°C of the protein solution was then measured for the wavelength range of 190 nm to 260 nm and the signal observed for buffer in the same cuvette was subtracted from these values to determine the CD signal for the protein. The CD data were then normalised by calculating the per residue mean residue ellipticity (MRE) using the following equation.

$$MRE = \frac{\theta}{10.N.c.l.}$$

where N is the number of amino acids in the protein of interest, c is the molar concentration of the protein and l is the pathlength of the cuvette in cm.

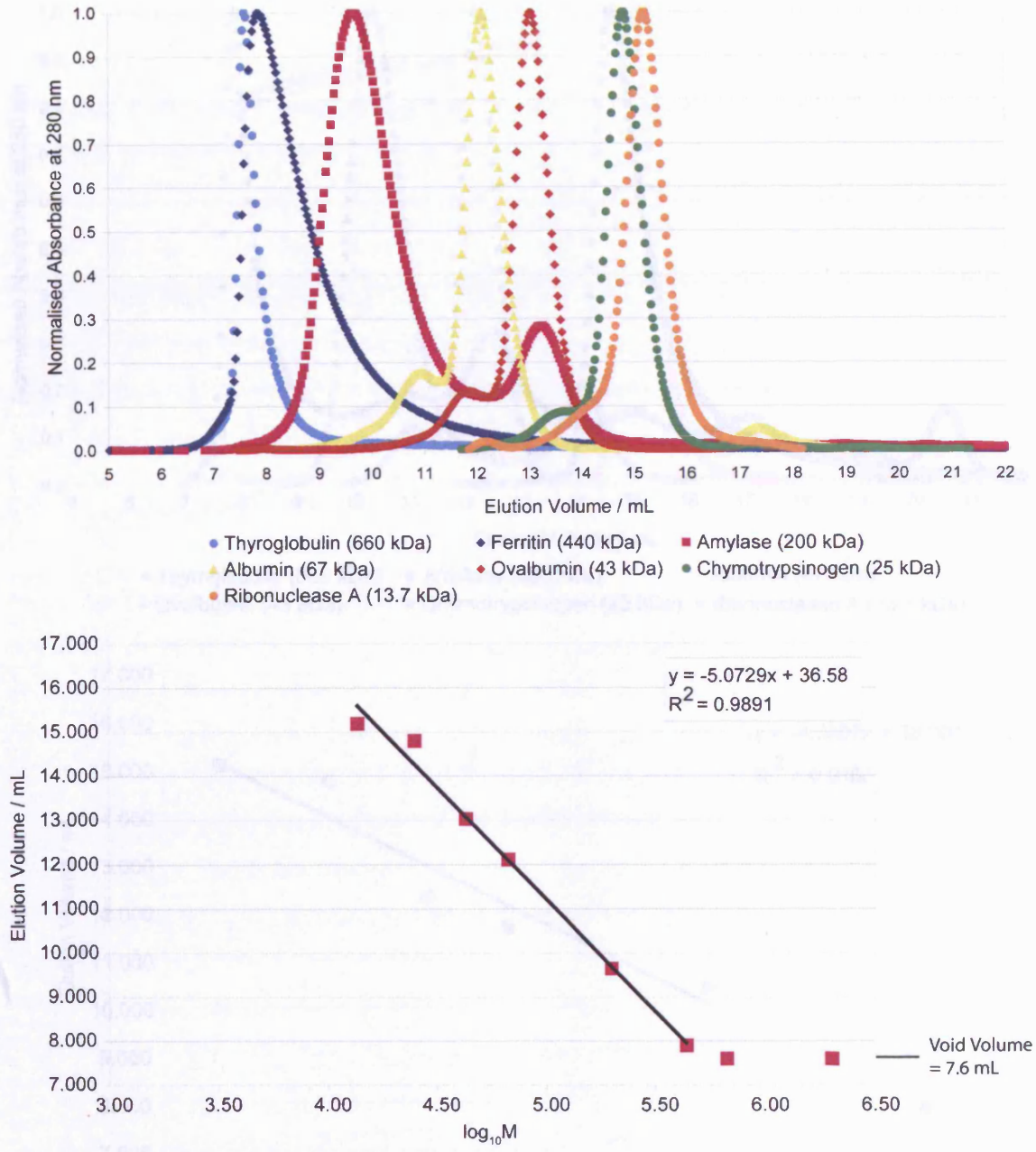


Figure 2-5. Calibration of the superose 12HR analytical size exclusion chromatography column in 20 mM potassium phosphate pH 7.0, 300 mM NaCl, 1 mM EDTA. The absorbance values for each protein elution profile have been normalised so that the highest value observed corresponds to an arbitrary absorbance of one absorbance unit. The calibration graph for the Superose 12HR column under these buffer conditions is shown below.

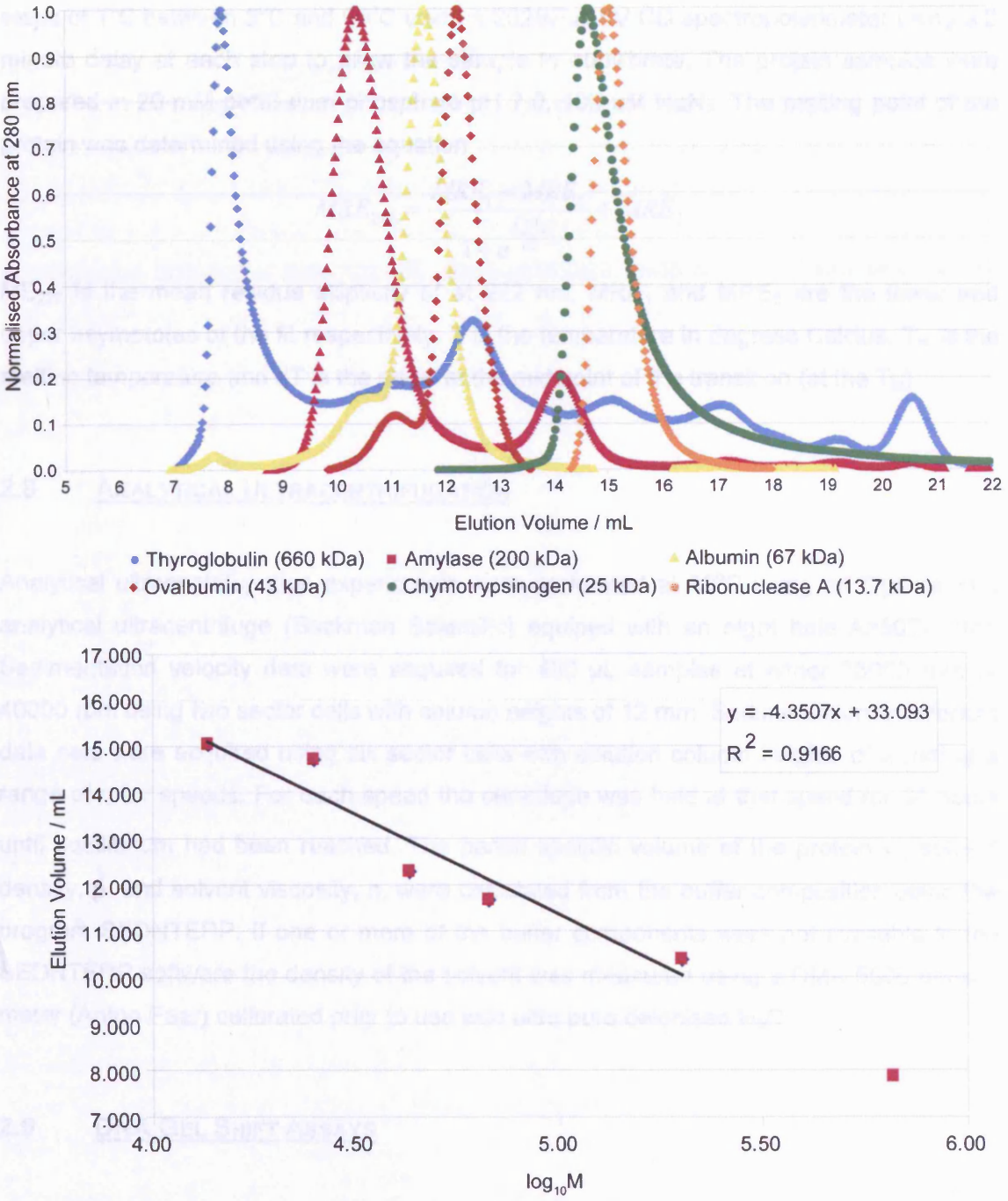


Figure 2-6 Calibration of the superose 12HR analytical size exclusion chromatography column in 20 mM potassium phosphate pH 7.0, 1 mM EDTA. The absorbance values for each protein elution profile have been normalised so that the highest value observed corresponds to an arbitrary absorbance of one absorbance unit. The calibration graph for the Superose 12HR column under these buffer conditions is shown below.

Melting curves for H-NS_{1-74} C21S were produced by monitoring the CD signal at 222 nm in steps of 1°C between 5°C and 95°C using a 202SF AVIV CD spectropolarimeter using a 2 minute delay at each step to allow the sample to equilibrate. The protein samples were prepared in 20 mM potassium phosphate pH 7.0, 100 μM NaN_3. The melting point of the protein was determined using the equation

$$MRE_{222} = \frac{MRE_1 - MRE_2}{1 + e^{\frac{T - T_m}{dT}}} + MRE_2$$

ME_{222} is the mean residue ellipticity of at 222 nm, MRE_1 and MRE_2 are the lower and upper asymptotes of the fit respectively, T is the temperature in degress Celcius, T_m is the melting temperature and dT is the slope at the mid point of the transition (at the T_m)

2.8 ANALYTICAL ULTRACENTRIFUGATION

Analytical ultracentrifugation experiments were performed at 20°C using an Optima XL-I analytical ultracentrifuge (Beckman Scientific) equiped with an eight hole An50Ti rotor. Sedimentation velocity data were acquired for 400 μL samples at either 35000 rpm or 40000 rpm using two sector cells with column heights of 12 mm. Sedimentation equilibrium data sets were acquired using six sector cells with solution column heights of 2 mm at a range of rotor speeds. For each speed the centrifuge was held at that speed for 30 hours until equilibrium had been reached. The partial specific volume of the protein, $\bar{\nu}$, solvent density, ρ, and solvent viscosity, η, were calculated from the buffer composition using the program SEDNTERP. If one or more of the buffer components were not available in the SEDNTERP software the density of the solvent was measured using a DMA 5000 density meter (Anton Paar) calibrated prior to use with ultra pure deionised H_2O.

2.9 DNA GEL SHIFT ASSAYS

2.9.1 Preparation of DNA.

To test the binding affinity and stoichiometry of the interaction of H-NS and StpA proteins for AT rich DNA sequences various DNA oligonucleotides were synthesised by MWG corrsponding to different lengths of DNA. Each oligonucleotide was based around a

repeating AAT sequence with a GC or CG sequence at the ends of the DNA to stabilise the ends of the DNA duplex. An AAT repeating sequence was chosen as H-NS preferentially binds to AT–rich sequences . The oligonucleotide sequences and their calculated molar extinction coefficients and melting tempeatures are described in table 5-1. Each oligonucleotide was dissolved in 10 mM sodium phosphate pH 7.0, 100 mM NaCl, 10 μM NaN_3 to a final DNA concentration of 200 μM. Equal amounts of the complementary oligonucleotides were mixed together and heated to 95°C in a boiling 500 mL water bath for 5 minutes. The water bath was then allowed to cool slowly to room temperature before being transfered into a 4°C refridgerator to complete the annealing process.

2.9.2 DNA Gel Shift Assays for dsDNA AAT Repeat Sequences.

10 μL solutions containing 2 μM double stranded DNA and various concentrations of H-NS_{FL}, H-NS_{90-137}, $StpA_{FL}$ or $StpA_{91-134}$ were dissolved in 10 mM sodium phosphate pH 7.0, 100 mM NaCl, 10 μM NaN_3 were incubated on ice for thirty minutes. After this incubation period 2 μL of 12% (w/v) ficoll type 400 was added. The solutions were then applied onto a 8% non-denaturing polyacrylamide gel in 45 mM Tris-borate pH 8.3, 1 mM EDTA at 4°C, 10 V cm^{-1}. A sample of O' Gene Ruler ultra low range DNA ladder (Fermentas) was applied to each gel to allow direct comparison of the DNA positions between gels. After staining the gel in a solution of 0.5 μg mL^{-1} ethidium bromide the gels were analysed using a UV transilluminator.

2.10 NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

All experiments were recorded at 25°C on Varian Unity INOVA or Varian Unity PLUS spectrometers operating at nominal ^1H frequencies of 600 MHz and 500 MHz respectively. NMR experiments were set up with assistance from Dr M. Williams or Dr R. Harris.

All NMR data processing was performed using NMRpipe (Delaglio et al., 1995). NMR spectra were analysed and the cross peak assignments determined using ANSIG 2.7 (Kraulis, 1989; Kraulis et al., 1994).

2.10.1 1D ^{1}H NMR

All one dimensional ^{1}H-NMR experiments were acquired using WATERGATE solvent suppression. The experiments were setup with the number of points (np) set to 8192 for a sweep width (sw) of 10000.0 Hz. Spectra were processed using NMRpipe (Delaglio et al., 1995) with a sine-bell squared apodisation function. 50 µM protein samples for H-NS_{1-68} C21S, H-NS_{1-71} C21S, H-NS_{1-74} C21S, H-NS_{1-77} C21S and H-NS_{1-83} C21S were prepared in 20 mM potassium phosphate pH 7.0, 300 mM NaCl, 1 mM NaN_3 and 10% D_2O.

2.10.2 2D [^{1}H,^{15}N]-HSQC

A 2D [^{1}H,^{15}N]-HSQC spectrum (Kay et al., 1992) was acquired on a 1 mM [^{1}H,^{15}N]-labelled sample of H-NS_{1-74} C21S containing 20 mM potassium phosphate, pH 7.0, 300 mM NaCl, 1 mM EDTA, 10 µM NaN_3 and 10% D_2O at 25°C using WATERGATE solvent supression. The experiment parameters are described in Table 2-6.

2D [^{1}H,^{15}N]-HSQC spectra were also acquired on 2 mM samples of [^{1}H,^{15}N]-labelled H-NS_{90-137} and $StpA_{91-134}$ containing 10 mM sodium phosphate pH 7.0, 100 mM NaCl, 10 µM NaN_3 at 25°C. The experiment parameters are described in Table 2-7.

All of the 2D [^{1}H,^{15}N]-HSQC spectra were processed with zero filling and a sine-bell squared apodisation function to both the ^{1}H and ^{15}N dimensions prior to fourier transformation.

2.10.3 (^{1}H)-^{15}N heteronuclear NOE

The (^{1}H)-^{15}N heteronuclear NOE experiment (Farrow et al., 1994) was acquired on a 1 mM [^{1}H,^{15}N]-labelled sample of H-NS_{1-74} C21S containing 20 mM potassium phosphate pH 7.0, 300 mM NaCl, 1 mM EDTA, 10 µM NaN_3. The (^{1}H)-^{15}N heteronuclear NOE data were recorded with 3.0 seconds ^{1}H saturation in the latter part of a 3.5 seconds preparation period delay, which was also used without radio frequency pulses for the reference 2D spectrum without NOE. The NMR parameters are described in Table 2-7. NMRpipe (Delaglio et al., 1995) was used to precess the results. Zero filling and a sine bell squared apodisation function was applied prior to fourier transformation.

Table 2-6. NMR spectroscopy parameters for 2D [^{1}H,^{15}N]-HSQC analysis of H-NS and StpA constructs.

Protein	^{1}H		^{15}N	
	np	sw	np	sw
H-NS$_{1-58}$ C21S	2048	8000	128	2000
H-NS$_{1-74}$ C21S	1024	4600.1	128	2000
H-NS$_{90-137}$	1024	4000	64	1600
StpA$_{91-137}$ (NaCl titration)	1024	4800	64	2000
StpA$_{91-137}$ (DNA titration)	1024	4000	64	1550

Table 2-7. NMR spectroscopy parameters for (^{1}H)-^{15}N-HNOE for H-NS$_{1-74}$ C21S.

^{1}H direct dimension		^{1}H indirect dimension	^{15}N	
np	sw	sw	np	sw
1792	6999	1000	128	1940

Table 2-8. NMR spectroscopy parameters for 3D ^{15}N-separated TOCSY-HSQC for H-NS$_{90-137}$.

^{1}H direct dimension		^{1}H indirect dimension		^{15}N		Mixing time/ ms
np	sw	np	sw	np	sw	
1024	8000	96	6000	64	1200	80

Table 2-9. NMR parameters for 3D ^{15}N-NOESY-HSQC for H-NS$_{90-137}$.

^{1}H direct dimension		^{1}H indirect dimension		^{15}N		Mixing time/ ms
np	sw	np	sw	np	sw	
1024	8000	96	6000	64	1200	120

2.10.4 ^{15}N-separated TOCSY-HSQC and ^{15}N-edited NOESY-HSQC

A ^{15}N-separated TOCSY-HSQC (Marion *et al.*, 1989) and ^{15}N-edited NOESY-HSQC (Bax *et al.*, 1990) was acquired on a 2 mM [^{1}H,^{15}N]-labelled sample of *S. typhimurium* H-NS$_{90-137}$ containing 10 mM sodium phosphate pH 7.0, 100 mM NaCl, 10 µM NaN_3. The NMR parameters for the ^{15}N-separated TOCSY-HSQC and ^{15}N-edited NOESY-HSQC are described in Table 2-8 and Table 2-9 respectively. The data matrices from NOESY and TOCSY experiments were processed with zero filling in both t2 and t1 domains. A Lorentz-Gaussian apodisation function was applied to the t1 time domains (LB = -4.0 Hz, GB = 0.01 Hz) prior to fourier transformation. A sine bell squared apodisation function was applied to the indirect dimensions prior to fourier transformation.

2.10.5 DNA Titration with H-NS$_{91-137}$ or StpA$_{91-134}$ by NMR Spectroscopy.

For the preparation of double stranded DNA (dsDNA) AAT20mer for the NMR titrations this DNA was further purified after annealing the two complementary oligonucleotides (AAT20mer forward and AAT20mer reverse) using a 20 mL Q-sepharose ion exchange chromatography column. The DNA was bound to the column after pre-equilibration with 10 mM sodium phosphate pH 7.0, 100 mM NaCl, 10 µM NaN_3 and eluted using a linear gradient from zero to 1 M NaCl run over 200 mL solvent volume. The fractions containing dsDNA AAT20mer were the pooled and dialysed against 11 mM sodium phosphate pH 7.0, 111 mM NaCl, 11µM NaN_3. D_2O was added to the DNA solution to a final concentration of 10% D_2O, reducing the other buffer components to 10 mM sodium phosphate, 100 mM NaCl, 10µM NaN_3. Stock solutions of 3040 µM and 3780 µM AAT20mer duplex DNA was prepared for the NMR titration with H-NS$_{90-137}$ and StpA$_{91-134}$ respectively.

500 µL NMR samples containing 300 µM *S. typhimurium* [^{1}H,^{15}N]-labelled H-NS$_{90-137}$ or *E. coli* [^{1}H,^{15}N]-labelled StpA$_{91-134}$, dissolved in 10 mM sodium phosphate pH 7.0, 100 mM NaCl, 10 µM NaN_3, 10% D_2O, were prepared. The volumes of the DNA solution added and the resulting concentrations of protein and DNA for each point in the titration are shown in Table 2-10 for H-NS$_{90-137}$ and Table 2-11 for StpA$_{91-134}$. At each point in the titration a [^{1}H,^{15}N]-HSQC experiment was acquired using WATERGATE solvent suppression. The experiment parameters used are described in Table 2-7

2.10.6 NaCl Titration with H-NS_{91-137} or $StpA_{91-134}$ by NMR Spectroscopy.

500 µL NMR samples containing 300 µM *S. typhimurium* ^{15}N-labelled H-NS_{90-137} or *E. coli* ^{15}N-labelled $StpA_{91-134}$, dissolved in 10 mM sodium phosphate pH 7.0, 100 mM NaCl, 10 µM NaN_3, 10% D_2O, were prepared. The volumes of a 10 mM sodium phosphate pH 7.0, 2 M NaCl, 10 µM NaN_3, 10% D_2O added to the sample and the resulting effect on the protein concentration and NaCl concentration in the sample are shown in Table 2-12 for H-NS_{90-137} and Table 2-13 for $StpA_{91-134}$. A [^{1}H,^{15}N]-HSQC spectrum was recorded for each point in the titration.

2.11 CRYSTALLISATION OF H-NS OLIGOMERISATION DOMAIN CONSTRUCTS

2.11.1 Crystallisation Trials

Crystallisation trials were performed using either a hanging drop vapour diffusion or microbatch setup. For the hanging drop vapour diffusion setup 0.5 mL of crystallisation condition was pipetted into the reservoir and 1 µL of the protein solution was pipetted into 1 µL of crystallisation condition solution, without mixing, on to a siliconised glass coverslip. The glass coverslip was then inverted and suspended in a sealed volume above the reservoir. For the microbatch setup 5 mL of paraffin oil was dispensed onto a Terasaki plate. Then 1 µL of protein solution was dispensed under the oil into each well and 1 µL of the crystallisation condition was added to this drop. The crystallisation trays were incubated at either 16°C or 20°C and examined regularly using a light microscope (Leica).

2.11.2 Testing X-ray Diffraction

Data were collected using a Rigaku RAXIS IV image-plate system and osmic focusing mirrors. Initial 15 minute exposure images were collected for 0.5° crystal rotation starting at 0° and 90° positions respectively. Data processing was performed using Crystal clear 1.3.6 SP3.3 software (Rigaku MSC)

Table 2-10. NMR titration of 3040 μM dsDNA AAT20mer into H-NS_{90-137}.

Total DNA Added/ μL	Total Volume/ μL	[Protein]/ μM	[DNA]/ μM	Molar Ratio DNA:Protein
0.0	500.0	300.0	0.0	0.000
0.3	500.3	299.9	1.5	0.005
0.5	500.5	299.7	3.0	0.010
2.0	502.0	298.8	12.1	0.041
4.0	504.0	297.6	24.1	0.081
6.0	506.0	296.4	36.1	0.122
8.0	508.0	295.3	47.9	0.162
10.0	510.0	294.1	59.6	0.203
12.0	512.0	293.0	71.3	0.243
14.0	514.0	291.8	82.9	0.284
16.0	516.0	290.7	94.3	0.324
18.0	518.0	289.6	105.7	0.365
20.0	520.0	288.5	117.0	0.406
22.0	522.0	287.4	128.2	0.446
25.0	525.0	285.7	144.9	0.507
28.0	528.0	284.1	161.3	0.568
33.0	533.0	281.4	188.3	0.669
38.0	538.0	278.8	214.9	0.771
48.0	548.0	273.7	266.5	0.973
58.0	558.0	268.8	316.2	1.176
78.0	578.0	259.5	410.5	1.582
108.0	608.0	246.7	540.4	2.190

Table 2-11. NMR titration of 3780μM dsDNA AAT20mer into $StpA_{91-134}$.

Total DNA Added/ μL	Total Volume/ μL	[Protein]/ μM	[DNA]/ μM	Molar Ratio DNA:Protein
0.0	500.0	300.0	0.0	0.000
0.5	500.5	299.7	3.8	0.013
2.0	502.0	298.8	15.1	0.050
5.0	505.0	297.0	37.4	0.126
8.0	508.0	295.3	59.5	0.202
11.0	511.0	293.5	81.4	0.277
14.0	514.0	291.8	103.0	0.353
17.0	517.0	290.1	124.3	0.428
20.0	520.0	288.5	145.4	0.504
23.0	523.0	286.8	166.3	0.580
26.0	526.0	285.2	186.9	0.655
29.0	529.0	283.6	207.3	0.731
32.0	532.0	282.0	227.4	0.807
35.0	535.0	280.4	247.3	0.882
38.0	538.0	278.8	267.0	0.958
41.0	541.0	277.3	286.5	1.033
46.0	546.0	274.7	318.5	1.159
51.0	551.0	272.2	349.9	1.285
61.0	561.0	267.4	411.1	1.537
66.0	566.0	265.0	440.9	1.664
71.0	571.0	262.7	470.1	1.790
76.0	576.0	260.4	498.8	1.916
81.0	581.0	258.2	527.1	2.042
91.0	591.0	253.8	582.1	2.294
101.0	601.0	249.6	635.4	2.546
111.0	611.0	245.5	686.8	2.798

Table 2-12. NMR titration of 10 mM sodium phosphate pH 7.0, 2 M NaCl, 10 μM NaN_3, 10% D_2O into H-NS_{90-137}.

Volume NaCl Solution Added/ μL	Total Volume/ μL	[Protein]/ μM	[NaCl]/ mM
0	500	300	100
10	510	294	139
20	520	288	177
30	530	283	208
40	540	278	248
60	560	268	314
80	580	259	376
100	600	250	433
120	620	242	468
140	640	234	538

Table 2-13. NMR titration of 10 mM sodium phosphate pH 7.0, 2 M NaCl, 10 μM NaN_3, 10% D_2O into $StpA_{91-134}$.

Volume NaCl Solution Added/ μL	Total Volume/ μL	[Protein]/ μM	[NaCl]/ μM
0.0	500.0	300.0	100.0
11.5	511.5	293.3	145.0
22.5	522.5	287.1	186.1
33.0	533.0	281.4	223.8
43.0	543.0	276.2	258.4
61.5	561.5	267.1	319.1
82.5	582.5	257.5	383.3
156.0	656.0	228.7	575.6
199.5	699.5	214.4	670.4
277.0	777.0	193.1	813.0
452.0	952.0	157.6	1050

For complete data collection the strategy software was used to determine the optimum crystal rotation for data collection. Data was collected for 65° of crystal rotation collecting an individual diffraction pattern image for every 1° of rotation of the crystal.

2.12 MODELLING THE STRUCTURE OF H-NS OLIGOMERISATION DOMAIN

The MODELLER 9v1 software package (Sali and Blundell, 1993; Sali et al., 1995) was used for comparative modelling of the H-NS oligomerisation domain.

3 CHARACTERISING THE HIGH ORDER OLIGOMERISATION INTERACTIONS OF H-NS

3.1 SUMMARY

One of the aims of this research was to gain a better understanding of the self-association interactions of H-NS, particularly those interactions required for high order oligomerisation (the formation of tetramers and larger oligomeric species). In this chapter the oligomeric properties of H-NS and various H-NS mutants are investigated along with the dependence of the oligomerisation interactions on ionic strength. This chapter describes the approach used to determine the amino acid sequence that comprises the complete oligomerisation domain of H-NS and shows that amino acids predicted to form a short α-helix at the C-terminal end of the oligomerisation domain are essential for the formation of tetramers and higher order oligomeric species. Sedimentation velocity and sedimentation equilibrium analytical ultracentrifugation (AUC) experiments demonstrate that the H-NS oligomerisation domain is able to form different oligomeric species, in a reversible concentration dependent manner and the effect of ionic strength on the oligomeric state is also investigated.

3.2 INTRODUCTION

Analysis of primary sequence alignments (Figure 1-5) of H-NS homologues identified two highly conserved domains; a N-terminal oligomerisation domain and a C-terminal nucleic acid binding domain (Bertin et al., 1999; Cusick and Belfort, 1998; Dorman et al., 1999; Tendeng and Bertin, 2003). The N-terminal 78 amino acids and the C-terminal amino acids, between Arg90 and Gln137 in *E. coli* H-NS, are well conserved whilst residues 79 to 90 show greater variation amongst H-NS homologues (Dorman et al., 1999). It has been previously shown that the N-terminal 90 amino acids are capable of forming high order oligomers but that the N-terminal 65 amino acids form a discrete homodimer (Badaut et al., 2002; Esposito et al., 2002). However, the structure of a complete oligomerisation domain has not been solved, so the interactions that are important for the high order oligomerisation of H-NS have yet to be determined. The two published NMR structures of H-NS_{2-47} (Bloch et al., 2003) and H-NS_{2-58} C21S (Esposito et al., 2002) provide evidence for some of the stabilising interactions within the coiled-coil region. However, the different relative orientation of the α-helices, in the

coiled-coil region (parallel versus antiparallel) between these two structures needs to be resolved, if the structure and how the structure of H-NS dictates its function are to be understood. It is important to note that in both of these structures the coiled-coil region, which is predicted to span from residues Leu23 to Arg62, is truncated. The truncation of the coiled-coil region could provide an explanation for the observed difference in the structures, with the shorter construct forming an antiparallel coiled-coil whereas a parallel coiled-coil configuration is observed for H-NS_{2-58} C21S. The secondary structure prediction algorithms: Porter (Pollastri and McLysaght, 2005), Prof (Ouali and King, 2000), SSpro8 (Pollastri et al., 2002), PsiPred (McGuffin et al., 2000) and Jpred (Cuff et al., 1998) all predict the third helix of H-NS to span from residues Leu23 to Ala67 (Figure 1-6). Additionally, the secondary structure predictions for H-NS identify a short fourth helix between residues Pro72 and Ala81 and the role of these amino acids in the formation of high order oligomers are investigated further in this chapter.

3.3 IDENTIFYING THE RESIDUES REQUIRED FOR HIGH ORDER OLIGOMERISATION

In order to determine which residues are required for high order oligomerisation the DNA encoding the N-terminal 83 amino acids of H-NS was amplified by PCR and cloned into a pET30A(tev) vector, between the NcoI and EcoRI restriction sites of the multiple cloning site. A pET14b plasmid with H-NS_{1-90} C21S cloned between the NdeI and BamHI restriction sites, provided by Dr A. Petrovic, was used as the template for the PCR reaction. Several truncation mutants were then prepared, by introducing stop codons at the appropriate positions in the DNA sequence. The stop codon terminates the translation of the messenger RNA, by the ribosome, resulting in the expression of C-terminally truncated forms of the protein: H-NS_{1-68} C21S, H-NS_{1-71} C21S, H-NS_{1-74} C21S, H-NS_{1-77} C21S, H-NS_{1-80} C21S or H-NS_{1-83} C21S. All of the oligomerisation domain constructs tested have a C21S mutation to avoid complexities of disulphide bond formation in biophysical experiments. This mutation does not affect the far-UV circular dichroism spectrum, melting characteristics or the DNA binding properties of H-NS (Smyth et al., 2000). Cys21 is not highly conserved among H-NS homologues (Figure 1-5). All of the protein constructs have an N-terminal histidine tag (His_6-tag) to facilitate the purification of the protein by metal affinity chromatography. After cleavage of the His_6-tag, by Tev protease, two additional amino acids (glycine and alanine), which are not present in the wild type H-NS protein, remain at the N-terminus of the protein.

3.3.1 Circular Dichroism of H-NS Constructs

Circular dichroism (CD) is a form of spectroscopy based on the differential absorption of left- and right-handed circularly polarised light by chiral molecules. The observed far UV spectrum (260 nm to 190 nm) can be used to assess the proportions of secondary structure in a protein because α-helix, β-sheet and random coil polypeptides give rise to characteristic circular dichroism spectra (Greenfield, 1996) in this wavelength range. CD was used to assess the foldedness of the constructs. All of the six truncated H-NS proteins give rise to far UV circular dichroism spectra with two observed minima at 208 nm and 222 nm, indicative of a predominantly α-helical structure (Figure 3-1). CD is very sensitive to protein concentration differences. The slight differences in signal intensity observed are likely to be due to errors in the protein concentration determination or small differences in the level of purity of the proteins prepared, as opposed to reflecting significant changes in the global protein structure.

3.3.2 Nuclear Magnetic Resonance Spectroscopy

Electrons, protons and neutrons have associated with them a quantum mechanical property called spin. In NMR, we are interested in the net overall spin of the nucleus (protons + neutrons) of an atom. This nuclear spin is quantised and can only have values of 0, ½, 1, 1½.... Nuclei with an odd number of protons, neutrons or both have a net nuclear spin but in protein NMR spectroscopy we are only concerned with nuclei with spin ½. Some common biologically relevant nuclei (^{12}C and ^{16}O) have nuclear spin of zero and do not give rise to NMR spectra. Additionally those nuclei that have a nuclear spin of ≥ 1 (such as ^{14}N) possess an electric quadrupole moment in addition to their magnetic moment. The quadrupole moment interacts with the electric field at the nucleus in such a way that there is a very efficient relaxation of the nuclear spin back to thermal equilibrium, resulting in the rapid decay of any NMR spectroscopy signal produced by these nuclei. In protein NMR spectroscopy the important nuclei are ^{1}H, ^{13}C and ^{15}N because these have a nuclear spin of ½ and give rise to detectable NMR signals. It is therefore necessary, depending on the NMR experiment being performed, to prepare uniformly labelled protein samples with ^{13}C carbon atoms and/or ^{15}N nitrogen atoms, instead of the more naturally occuring ^{12}C and ^{14}N isotopes.

When these nuclei are placed in an externally applied magnetic field (B_0), their magnetic moments align either parallel or anti-parallel with the applied field. It is

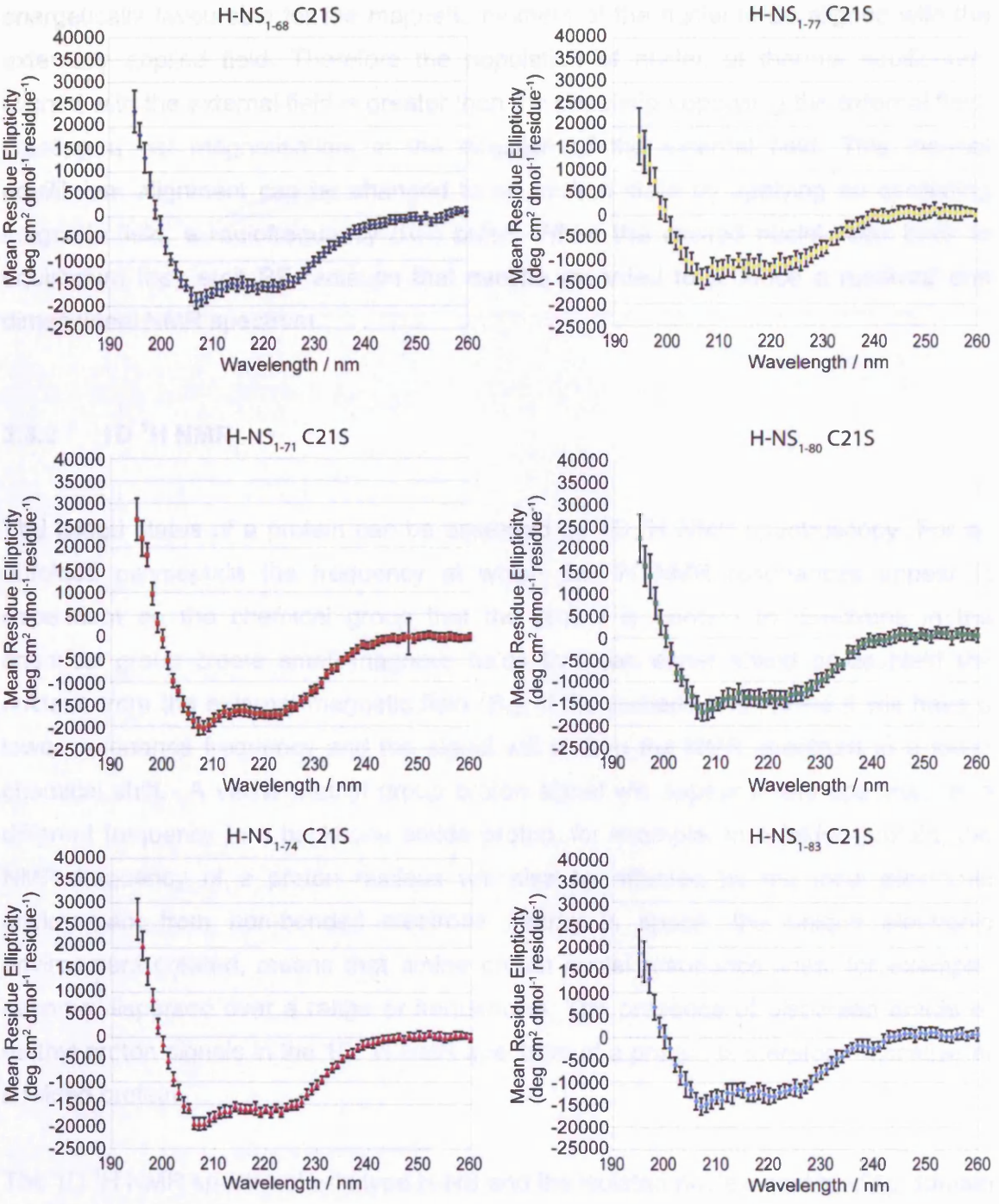


Figure 3-1. Circular dichroism spectra of H-NS oligomerisation domain constructs. Samples containing 50 μM protein in 20 mM potassium phosphate, pH7.0, 300 mM NaCl, were prepared. Spectra between 260 nm and 195 nm were recorded. Error bars show the variation of the circular dichroism signal observed over a five second averaging time for each wavelength.

energetically favourable for the magnetic moment of the nuclei to be aligned with the externally applied field. Therefore the population of nuclei, at thermal equilibrium, aligned with the external field is greater than the population opposing the external field, creating a net magnetisation, in the direction of the external field. This thermal equilibrium alignment can be changed to an excited state by applying an oscillating magnetic field, a radiofrequency (RF) pulse. When the excited nuclei relax back to equilibrium they emit RF radiation that can be recorded to produce a resolved one dimensional NMR spectrum.

3.3.3 1D ^{1}H NMR

The folded status of a protein can be assessed by 1D ^{1}H NMR spectroscopy. For an unfolded polypeptide the frequency at which the ^{1}H NMR resonances appear is dependent on the chemical group that the proton is bonded to. Electrons in the chemical group create small magnetic fields that can either shield or deshield the nucleus from the external magnetic field (B_0). If the nucleus is shielded it will have a lower resonance frequency and the signal will shift in the NMR spectrum to a lower chemical shift. A valine methyl group proton signal will appear in the spectrum at a different frequency to a backbone amide proton, for example. In a folded protein, the NMR frequency of a proton nucleus will also by affected by the local electronic environment from non-bonded electrons nearby in space; the unique electronic environment created, means that amide proton nuclei resonance lines, for example, become dispersed over a range of frequencies. The presence of dispersed amide or methyl proton signals in the 1D ^{1}H NMR spectrum of a protein is therefore indicative of a folded protein.

The 1D ^{1}H NMR spectra of wild type H-NS and the isolated nucleic acid binding domain of H-NS alone are almost identical to each other (Smyth et al., 2000). Additionally, in the DQF-COSY NMR spectra of wild type H-NS, CαH-NH COSY cross peaks, corresponding to scalar coupling (J-coupling) between the backbone amide proton and the carbon alpha proton of the same amino acid, are visible for amino acids found within the C-terminal nucleic acid binding domain (Shindo et al., 1995). However, there are only a few visible cross peaks in the DQF-COSY spectra for the other amino acids in the protein (residues 1 to 90). These experiments suggest that the N-terminal domain is either too large, or in intermediate exchange between multiple states, so that it cannot be seen using these NMR experiments. However, the nucleic acid binding

domain is able to act independently to the N-terminal oligomeisation domain and is visible in these NMR spectra.

1D ^{1}H NMR spectroscopy was used to assess the degree of folding of the truncated proteins at 25°C, in 20 mM potassium phosphate, pH 7.0, 300 mM NaCl. The spectra can be directly compared as the same protein concentration (50 μM) was used for each experiment. Spectra were processed using NMRpipe (Delaglio et al., 1995). The regions of the spectra corresponding to amide proton chemical shifts (6 to 10 ppm) and dispersed methyl group proton chemical shifts (-0.5 to 0.5 ppm) are shown in (Figure 3-2). The 1D ^{1}H spectra of H-NS_{1-68} C21S, H-NS_{1-71} C21S and H-NS_{1-74} C21S are characteristic of a folded protein, with dispersed proton peaks (between 8.5 ppm and 10 ppm) in the amide region of the spectra. However, there are very few dispersed amide and methyl peaks in the 1D ^{1}H NMR spectrum of these proteins because there is only one aromatic amino acid in the protein (Tyr61).

The spectra for the longer constructs: H-NS_{1-77} C21S, H-NS_{1-80} C21S and H-NS_{1-83} C21S however, show significantly less signal intensity relative to the spectra of H-NS_{1-74} C21S. There are no dispersed proton peaks in either the amide or methyl regions of the NMR spectra. A lack of dispersed proton peaks is often indicative of an unfolded protein but this would usually give rise to sharp NMR signals around 8 ppm. Instead the 1D ^{1}H NMR spectra of H-NS_{1-77} C21S, H-NS_{1-80} C21S and H-NS_{1-83} C21S show that the constructs form a larger oligomeric complex than the shorter H-NS_{1-74} C21S protein, under these experimental conditions.

Larger proteins have longer rotational correlation times and subsequently shorter transverse relaxation (T_2) times, resulting in a more rapid decay of the NMR signal observed (Evans, 1996). Most NMR experiments are limited to proteins less than 30 kDa in size because the rapid transverse relaxation of NMR signals for larger species results in both a loss of observed signal intensity and resolution between the different peaks in the spectra.

3.3.4 Analytical Size Exclusion Chromatography of H-NS Constructs

Analytical size exclusion chromatography (SEC) is a technique that is widely used to assess the apparent size of a protein in solution, as well as the heterogeneity of the

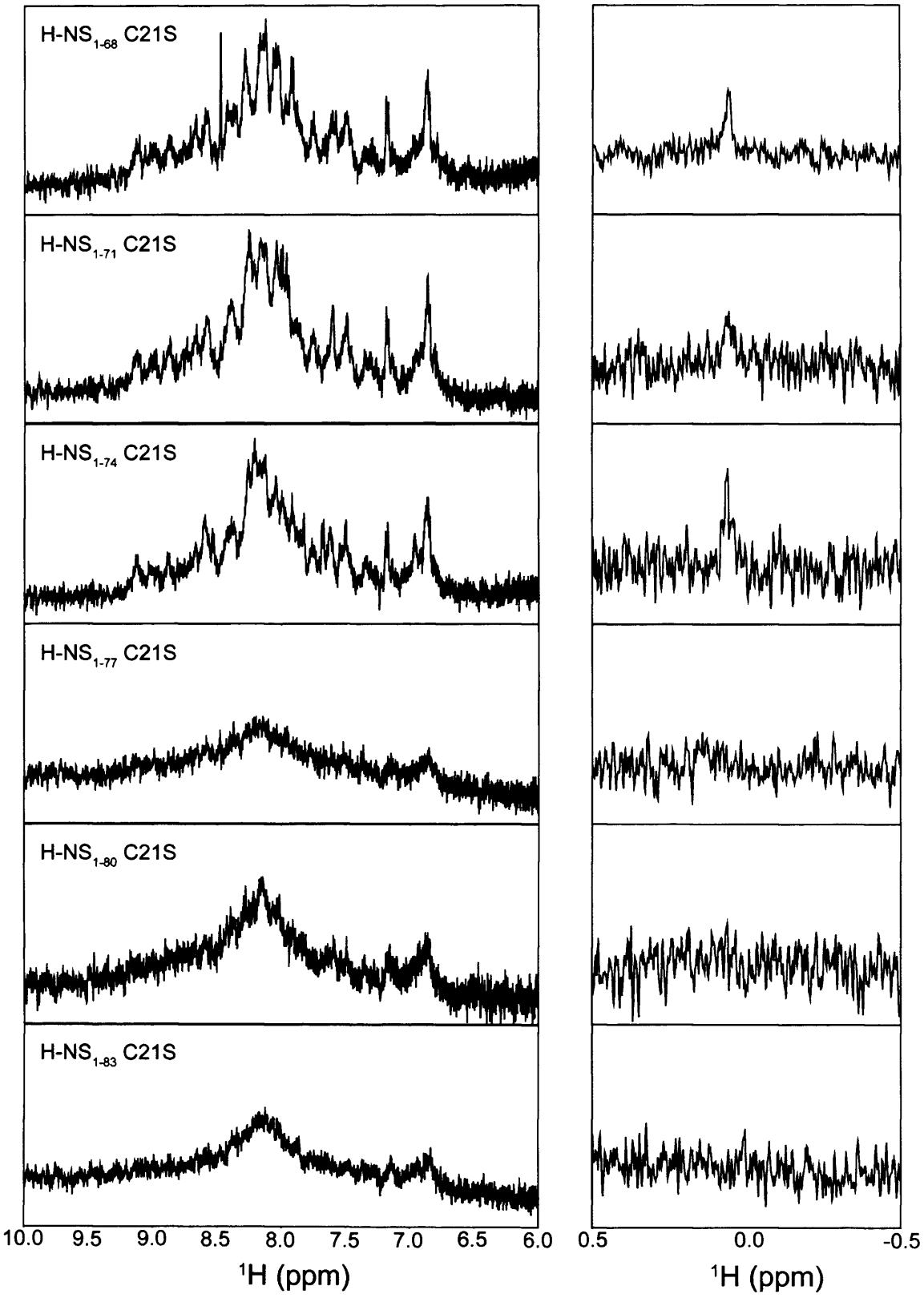


Figure 3-2 ^{1}H NMR spectra for C-terminally truncated H-NS proteins. The figure shows the ^{1}H NMR spectra for the amide region on the left (6.0ppm to 10.0ppm) and part of the methyl region where dispersed methyl signals are observed (-0.5ppm to 0.5ppm). The pictures have been scaled to give comparable baseline noise to allow direct comparison of the signal intensities observed.

protein sample. Most SEC studies use a zonal gel chromatography method where a small volume of solute is applied to a column equilibrated with buffer. For the study of proteins the column is packed with agarose, dextran or acrylamide based resin that contains a range of pore sizes, depending on the level of cross-linking between the polymer chains, which filter the proteins according to their Stokes radii. The larger the Stokes radius of the protein the fewer pores the protein can fit into and therefore the faster it will pass through the column. Smaller proteins become retarded for longer by the column and will elute at a later elution volume. As the zone of solute passes through the column it will be diluted, due to axial dispersion, so the eluted protein will be less concentrated than the applied solute concentration (Winzor, 2003). For a solute system comprising protein monomer in rapidly established equilibrium with higher order oligomeric species, a change in solute concentration gives rise to a change in the proportions of monomer and higher order oligomeric species present. The dilution effect observed as the solute passes through the analytical SEC column favours the dissociation of high order oligomeric species.

The oligomerisation properties of H-NS constructs were assessed by analytical SEC using a Superose 12HR column suitable for resolving globular proteins of molecular weights from 1 kDa to 300 kDa. The protein samples were dissolved in 20 mM potassium phosphate, pH 7.0, 300 mM NaCl, 1 mM EDTA. 100 µL protein samples at the concentrations shown were passed through the analytical size exclusion chromatography column at a flow rate of 0.75 mL min^{-1}. (Figure 3-3).

H-NS_{1-74} C21S elutes from the SEC column at approximately the same elution volume (~13 mL) for each protein concentration tested. The profile of the protein peaks observed is pseudo-symmetrical suggesting that the protein sample is homogeneous. There is a slight shift towards an earlier elution volume, as the protein concentration increases, as observed previously for H-NS_{1-65} C21S (Smyth et al., 2000; Badaut et al., 2002). There is also a small peak at approximately 11.9 mL which could indicate the presence of a larger species at higher concentrations. When a freshly prepared sample of 3.7 mM H-NS_{1-74} C21S was passed through this SEC column a single peak with elution volume of 12.43 mL was observed (Figure 3-3 inset) suggesting that this peak, at 11.9 mL is due to aggregation of the protein during storage at 4°C.

In comparison to the results for H-NS_{1-74} C21S, when increasing concentrations of H-NS_{1-77} C21S and H-NS_{1-83} C21S were passed through the same SEC column there is a significant shift in the profile of the protein peak towards smaller elution volumes as the

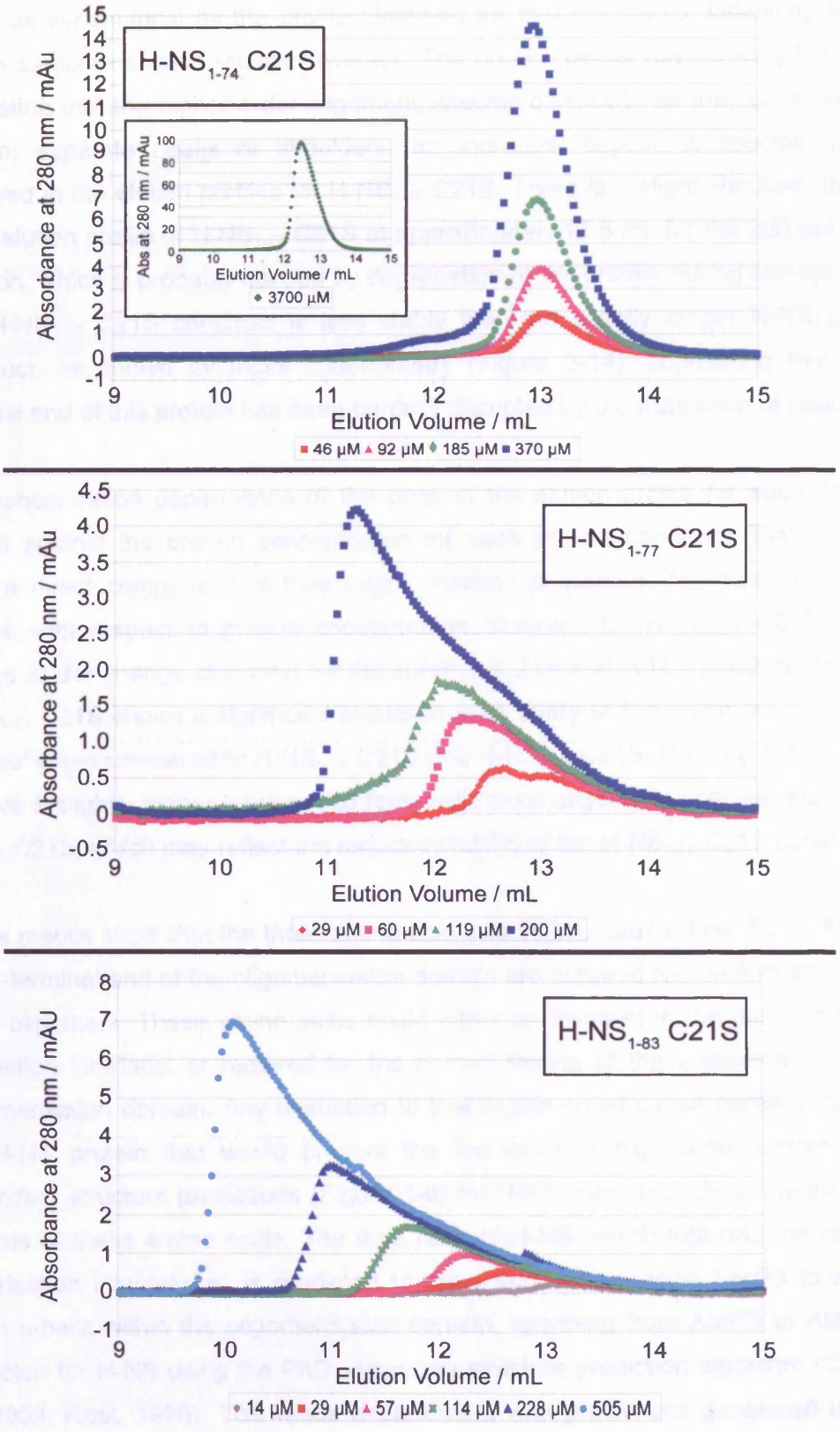


Figure 3-3 Analytical size exclusion chromatography titration of H-NS_{1-74} C21S, H-NS_{1-77} C21S and H-NS_{1-83} C21S. Inset in the H-NS_{1-74} C21S analytical SEC titration graphs is the absorption trace observed after injection of a 3700 µM H-NS_{1-74} C21S protein sample into the Superose 12HR column.

sample concentration is increased. In addition, the profile of the elution peak is no longer as symmetrical as the profile observed for H-NS_{1-74} C21S, indicating that the protein sample contains multiple species. The protein peaks have a long tail region, suggesting that the higher order oligomeric species dissociate as they pass down the column; separate peaks or shoulders, for individual oligomeric species, are not observed in the elution profiles for H-NS_{1-83} C21S. There is a slight shoulder observed in the elution profile of H-NS_{1-77} C21S at approximately 12.5 mL for the 200 μM protein injection, which is probably caused by degradation of the protein during storage at 4°C. The H-NS_{1-77} C21S construct is less stable than the slightly longer H-NS_{1-83} C21S construct, as shown by mass spectrometry (Figure 2-14); suggesting that the C-terminal end of this protein has been partially disrupted by the truncation at residue 77.

The concentration dependence of the peak of the elution profile for each protein is plotted against the protein concentration for each protein construct (Figure 3-4), to allow a direct comparison of their oligomerisation properties. The change in elution volume with respect to protein concentration observed for H-NS_{1-74} C21S is small relative to the change observed for the H-NS_{1-77} C21S and H-NS_{1-83} C21S constructs. H-NS_{1-74} C21S shows a significant reduction in its ability to form high order oligomeric species' when compared to H-NS_{1-77} C21S and H-NS_{1-83} C21S. H-NS_{1-83} C21S appears to have a slightly higher tendency to form high order oligomeric species relative to H-NS_{1-77} C21S, which may reflect the reduced stability of the H-NS_{1-77} C21S construct.

These results show that the three amino acid sidechains, Leu75, Leu76 and Asn77, at the C-terminal end of the oligomerisation domain are required for the formation of high order oligomers. These amino acids could either be involved in the dimer-to-tetramer interaction interface, or required for the correct folding of the C-terminal end of the oligomerisation domain. Any disruption to that region could cause partial unfolding of the H-NS protein that would prevent the formation of high order oligomers. The secondary structure predictions (Figure 1-6) for H-NS could provide some insight into the role of these amino acids. The third helix of H-NS, which includes the coiled-coil dimerisation interactions, is predicted to span from amino acids Leu23 to Ala67. A fourth α-helix within the oligomerisation domain, spanning from Asn73 to Ala82, was predicted for H-NS using the PhD secondary structure prediction algorithm (Cerdan et al., 2003; Rost, 1996). This result is consistent with predictions generated using the improved secondary structure prediction programs Porter (Pollastri and McLysaght, 2005), Prof (Ouali and King, 2000), PsiPred (McGuffin et al., 2000), SSpro8 (Pollastri et al., 2002), and Jpred (Cuff *et al.*, 1998) (Figure 1-6). These prediction algorithms suggest that a fourth α-helix starts at either Pro72 or Asn73 and ends

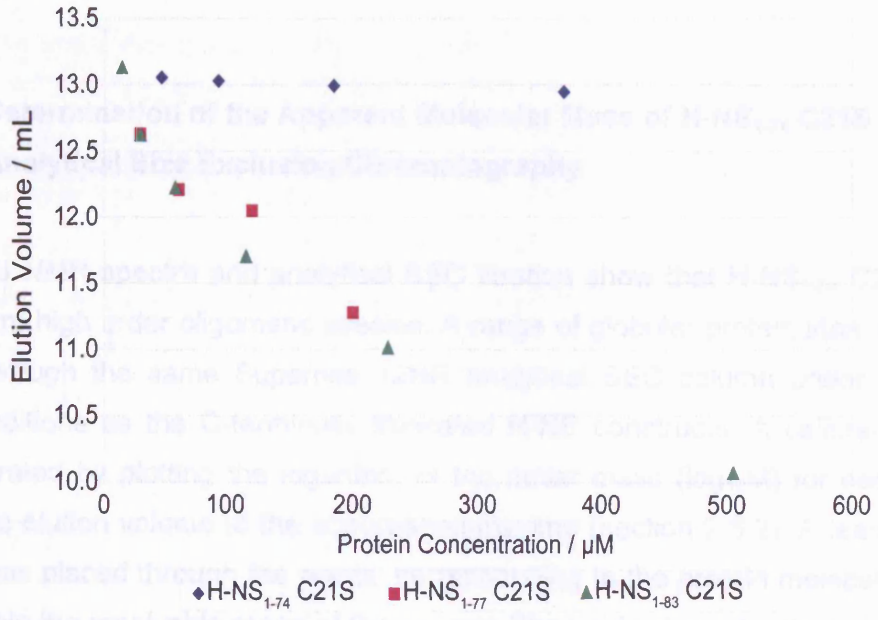


Figure 3-4. The loading concentration dependence of the elution profile maxima for H-NS_{1-74} C21S, H-NS_{1-77} C21S and H-NS_{1-83} C21S.

between Ser78 and Ala82. The 1D ^{1}H NMR spectra and analytical SEC results show that the amino acids in the region of the predicted fourth α-helix are required for the formation of larger oligomeric species.

3.4 EXAMINING THE STRUCTURAL PROPERTIES OF H-NS$_{1-74}$ C21S

3.4.1 Determination of the Apparent Molecular Mass of H-NS$_{1-74}$ C21S by Analytical Size Exclusion Chromatography

The 1D ^{1}H NMR spectra and analytical SEC titration show that H-NS$_{1-74}$ C21S is not able to form high order oligomeric species. A range of globular protein standards were passed through the same Superose 12HR analytical SEC column under the same buffer conditions as the C-terminally truncated H-NS constructs. A calibration graph was generated by plotting the logarithm of the molar mass ($\log_{10}$M) for each protein against the elution volume at the absorbance maxima (section 2.6.2). A least squares linear fit was placed through the points, corresponding to the protein molecules whose size is within the resolvable range of the column. The void volume for the column is 7.6 mL. Based on this calibration curve the apparent molar mass of H-NS$_{1-74}$ C21S is between 43 kDa and 46 kDa for the concentration range tested. This is approximately 5 times greater than the monomer molar mass of this protein (8.9 kDa). Previous studies of H-NS$_{1-65}$ C21S found that the protein eluted from the same analytical SEC column at a size corresponding to a trimer (Smyth et al., 2000). However sedimentation equilibrium analytical ultracentrifugation (AUC) experiments, a technique that directly measures the buoyant molar mass of a protein, in a shape independent manner, showed that H-NS$_{1-65}$ forms a stable homodimer in solution (Badaut et al., 2002; Esposito et al., 2002). The discrepancy between the results from these two techniques can be explained if H-NS has an elongated, as opposed to spherical, protein shape.

In a SEC experiment, the volume where a protein elutes, Ve, is dependant on the Stokes radius, a, of the protein (Ackers, 1967). The Stokes radius is the radius of the equivalent hydrodynamic sphere. The molecular mass, M, and Stokes radius of a solute are related to each other by the expression

$$a = \left(\frac{f}{f_0}\right)\left[\frac{3\bar{\nu}M}{4\pi N}\right]^{1/3}$$

where f/f_o is the frictional ratio, $\bar{\nu}$ is the partial specific volume of the solute and N is Avogadro's number. The use of a protein calibration curve, to determine the molecular

mass of a protein, requires the assumption that the frictional ratio and partial specific volumes of the proteins used are consistent for all of the proteins tested. AUC experiments using solvents of different densities demonstrate that the partial specific volume of proteins lie within the range of 0.70-0.75 mL g^{-1}, unless the protein is glycosylated; a reasonable assumption can be made that this parameter is constant for the proteins used to calibrate the column as well as the H-NS constructs (Winzor, 2003). The frictional ratio is the ratio of the frictional coefficient of the solvated solute molecule, f, relative to that of a sphere of the same volume as the unsolvated solute. Both the shape and degree of solvation can vary significantly between proteins, so the assumption that the frictional ratio is constant for all of the proteins passed through the SEC column can lead to significant errors in the determination of apparent molar masses. An elongated protein has a larger frictional ratio than a globular protein resulting in an over estimation of the apparent molar mass of the protein, if a globular protein calibration curve is used.

3.5 ANALYTICAL ULTRACENTRIFUGATION THEORY

In order to determine the actual size of H-NS_{1-74} C21S, various concentrations were examined by sedimentation equilibrium AUC. Unlike SEC, where the apparent molar mass of a protein is dependant on the Stokes radius of the protein, sedimentation equilibrium AUC can be used to directly determine the molar mass of a protein. Before the experimental results for the AUC analysis of H-NS_{1-74} C21S are discussed, the principals governing AUC experiments will be briefly explained.

The analytical ultracentrifuge can be considered as a conventional centrifuge equipped with an optical detection system, which allows the distribution of the protein in the sample chamber to be observed. The optical detection system is synchronised with the speed of the rotor, so that the data is acquired at the point when the sample is in the light path. The analytical ultracentrifuge used for this thesis is equipped with a UV visible spectrophotometer for absorption measurements, and a laser interferometer that records the refractive index gradient of the sample. Two different types of analytical ultracentrifugation techniques have been used in this research: sedimentation velocity AUC and sedimentation equilibrium AUC experiments.

3.5.1 Sedimentation Velocity AUC Experiments

In a sedimentation velocity experiment, the solute experiences a centrifugal force pulling it towards the bottom of the cell. There is a depletion of the macromolecules at the meniscus and the formation of a concentration boundary that moves towards the bottom of the centrifuge cell, as a function of time. The sedimentation coefficient, s, is a constant that describes the velocity, u, of the solute molecule per unit centrifugal force, $\omega^2 r$. The definition of the sedimentation coefficient is given by the Svedberg equation

$$s = \frac{u}{\omega^2 r} = \frac{M\left(1-\bar{\nu}\rho\right)}{Nf} = \frac{MD\left(1-\bar{\nu}\rho\right)}{RT}$$

M is the molar mass of the solute. $\bar{\nu}$ is the partial specific volume of the solute, ρ is the solvent density, N is Avogadro's number, f is the frictional coefficient, D is the diffusion coefficient, R is the molar gas constant and T is the temperature in Kelvin.

In a sedimentation velocity AUC experiment, the sedimentation of the protein is monitored over time. The sedimentation coefficient can be determined from the migration of the sedimentation boundary over time whilst the diffusion coefficient dictates the spread of the sedimentation boundary with time.

For a simple system the frictional coefficient, calculated from the experimentally determined sedimentation coefficient and diffusion coefficient, can be compared to that of a compact sphere of equal mass and density to the unsolvated solute, in order to determine the frictional ratio, an indicator of the low resolution shape of the molecule.

3.5.2 Modelling with Solutions of the Lamm Equation

The Lamm equation describes the change over time of the concentration distribution of the macromolecule in the centrifuge cell. This equation describes both the sedimentation and diffusion processes occurring inside the sample cell of the analytical ultracentrifuge.

$$\left(\frac{\delta c}{\delta t}\right)_r = -\frac{1}{r}\left(\frac{\delta}{\delta r}\left[s\omega^2 r^2 c - Dr\left(\frac{\delta c}{\delta r}\right)_t\right]\right)_t$$

$(\delta c/\delta t)_r$ is the change in solute concentration, c, at a radial distance from the centre of rotation, r, over a period of time, t (Svedberg and Pedersen, 1940). Direct fitting of the Lamm equation for individual species is, in practice, unreliable due to impurities and sample heterogeneity leading to broadening of the sedimentation boundary (Lebowitz

et al., 2002). In these circumstances, the diffusion coefficient is overestimated leading to an underestimation of the molecular mass of the macromolecule.

The c(s) distribution analysis (Dam and Schuck, 2004), implemented in the Sedphat software, calculates concentration distributions of single non-interacting species $\chi(s, D(s), r, t)$, using the Lamm equation for a large number of sedimentation coefficients ranging from s_{min} to s_{max} (Schuck, 2000), to find the best fit for the experimental data.

$$a(r,t) \cong \int_{s_{min}}^{s_{max}} c(s)\chi(s, D(s), r, t)ds$$

The hydrodynamic frictional ratio (f/f_0) for all of the macromolecule species in the sample is assumed to be constant. Therefore, the diffusion coefficient can be calculated for each sedimentation coefficient using the following relationship.

$$D(s) = \frac{\sqrt{2}}{18\rho} kTs^{-\frac{1}{2}} \left(h\left(\frac{f}{f_0}\right)\right)^{-\frac{1}{3}} \left[\frac{1-\bar{\nu}\rho}{\bar{\nu}}\right]^{-\frac{1}{2}}$$

f/f_0, in this case, is the weight averaged frictional ratio for the macromolecules present. The value for the weight averaged frictional ratio that gives the best agreement with the experimental data is determined by non linear regression. This approach relies on the weak dependence of the frictional ratio for the shape of the macromolecule, ranging from a value of 1.2 for globular proteins to around 1.8 for asymmetric proteins.

For a protein sample with multiple non-interacting species peaks will be observed in the c(s) distribution corresponding to the sedimentation coefficient of the various species present. Integrating the peak will determine the concentration of the protein present in the sample, if the molar extinction coefficient of the protein is known.

Although the c(s) analysis is based on the superposition of non-interacting species it can be a useful tool for the detection of interactions. Interactions can be detected by the emergence of new peaks at higher concentrations or a shift in the ratio of the peak areas with respect to loading concentration. Peaks that remain at a constant sedimentation coefficient but change in their relative area indicate the formation of complexes that are stable on the time-scale of sedimentation. The kinetics of association and dissociation of the complexes are slow so the species are not in equilbrium exchange during the experiment. However, a shift in the position of the peak indicates more rapid interconversion of species during sedimentation. For example, consider a protein that is in equilibrium between a monomer and a dimer. As the sedimentation velocity AUC experiment proceeds the monomeric and dimeric species

become separated with the larger dimeric complex sedimenting faster than the smaller monomer species. Dissociation of the protein dimer into the individual monomer components can occur if the off rate (k_{off}) is sufficiently fast, resulting in a broadened peak in the distribution that is present between the sedimentation coefficient of the dimer and the sedimentation coefficient of the monomer protein (Schuck, 1998; Schuck, 2003). The position of this peak reflects the relative proportions of monomer and dimer that are present in the sample.

3.5.3 Sedimentation Equilibrium Analytical Ultracentrifugation

In a sedimentation equilibrium analytical ultracentrifugation experiment, the protein sample experiences a relatively low centrifugal force (relative to sedimentation velocity experiments where higher rotor speeds are used). As the sedimentation process proceeds, the solute concentration at the bottom of the cell increases over time. Under these conditions an equilibrium becomes established where the sedimentation of the protein, due to the centrifugal force, is balanced by the diffusion of the protein back up the cell, due to the concentration gradient. For a single species under ideal conditions, where there is an absence of repulsive interactions, due to charge interactions or volume exclusion, the concentration distribution is described by the equation

$$a(r,t) = c(r_0)\varepsilon d.\exp\left[M\left(1-\bar{\nu}\rho\right)\frac{\omega^2}{2RT}\left(r^2 - r_0{}^2\right)\right]$$

a(r,t) is the absorbance observed at a radius, r, at time, t. $c(r_0)$ is the concentration of the solute at a reference radius, r_0, ε is the molar extinction coefficient of the solute, d is the optical pathlength and R is the molar gas constant. $M(1-\bar{\nu}\rho)$ is the buoyant molar mass of the protein which can be converted into the actual molar mass of the protein if the partial specific volume of the solute, $\bar{\nu}$, and the solvent density, ρ, is known. In a typical sedimentation equilibrium AUC experiment it takes between 24 and 48 hours to establish the equilibrium distribution for each rotor speed. Multiple rotor speeds are required to determine the mass and distribution of the species present.

3.5.4 Self-association in Sedimentation Equilibrium AUC Experiments

Models for the self-association of solute molecules can be used to interpret the binding constants for an interaction, if multiple loading concentrations and rotor speeds are used. If a protein self associates in a reversible manner, an equilibrium distribution will be reached where both a mechanical equilibrium, balancing the sedimentation and

diffusion of the solute molecules, and a chemical equilibrium, where the solute molecules of different oligomeric states distribute themselves according to the law of mass action. For a monomer-dimer equilibrium the amount of dimer present can be expressed as $c_2=K_{12}.c_1^2$ where K_{12} is the association constant for the interaction, c_2 is the concentration of dimer and c_1 is the monomer concentration. Therefore the sedimentation equilibrium profile follows the relationship

$$a(r) = c_1.(r_0).\varepsilon_1.d.\exp\left[M_b \frac{\omega^2}{2RT}\left(r^2 - r_0^2\right)\right] + K_{12}.c_1.(r_0)^2.2\varepsilon_1.d.\exp\left[2M_b \frac{\omega^2}{2RT}\left(r^2 - r_0^2\right)\right]$$

where M_b is the buoyant molar mass of the monomer, $M_b = (1-\bar{\nu}\rho)$. The concentration gradient established within the sample cell, once the equilibrium is reached, allows a wide range of concentrations to be studied if multiple rotor speeds and loading concentrations are used.

3.6 H-NS$_{1-74}$ C21S FORMS A HOMODIMER IN SOLUTION

Sedimentation equilibrium AUC analysis was performed at three rotor speeds with multiple loading concentrations of H-NS$_{1-74}$ C21S. Using the sample distribution profiles, once an equilibrium had been established, the oligomeric state of the protein could be assessed. Four different loading concentrations from 48 μM to 271 μM H-NS$_{1-74}$ C21S were used for the analysis. It should be noted that once the equilibrium distribution of the protein was established the concentration range examined was between 35 μM and 450 μM H-NS$_{1-74}$ C21S (the absorbance at 280 nm ranges from 0.05 absorbance units to 0.7 absorbance units) (Figure 3-5).

The absorbance at 280 nm was monitored and profiles for three rotor speeds, 35000 rpm, 38000 rpm and 40000 rpm showed sufficient curvature to allow for accurate fitting of the buoyant molar mass. Using the solvent density and partial specific volume of the protein, calculated using the software SEDNTERP (Laue et al., 1992), the buoyant molar mass observed, could be converted into the actual molar mass of the protein. A good fit was obtained for a global fit, of all twelve curves, using a model for a single species with a molar mass of 17459.6 g mol^{-1} (Figure 3-5). This molar mass corresponds to that of a homodimer of H-NS$_{1-74}$ C21S. The raw data could not be fitted to a single species model, when the molar mass was fixed to that of either a monomer (8729.8 g mol^{-1}), or a trimer (26189.4 g mol^{-1}) for H-NS$_{1-74}$ C21S. H-NS$_{1-74}$ C21S is restricted to being a homodimer in solution under the conditions tested.

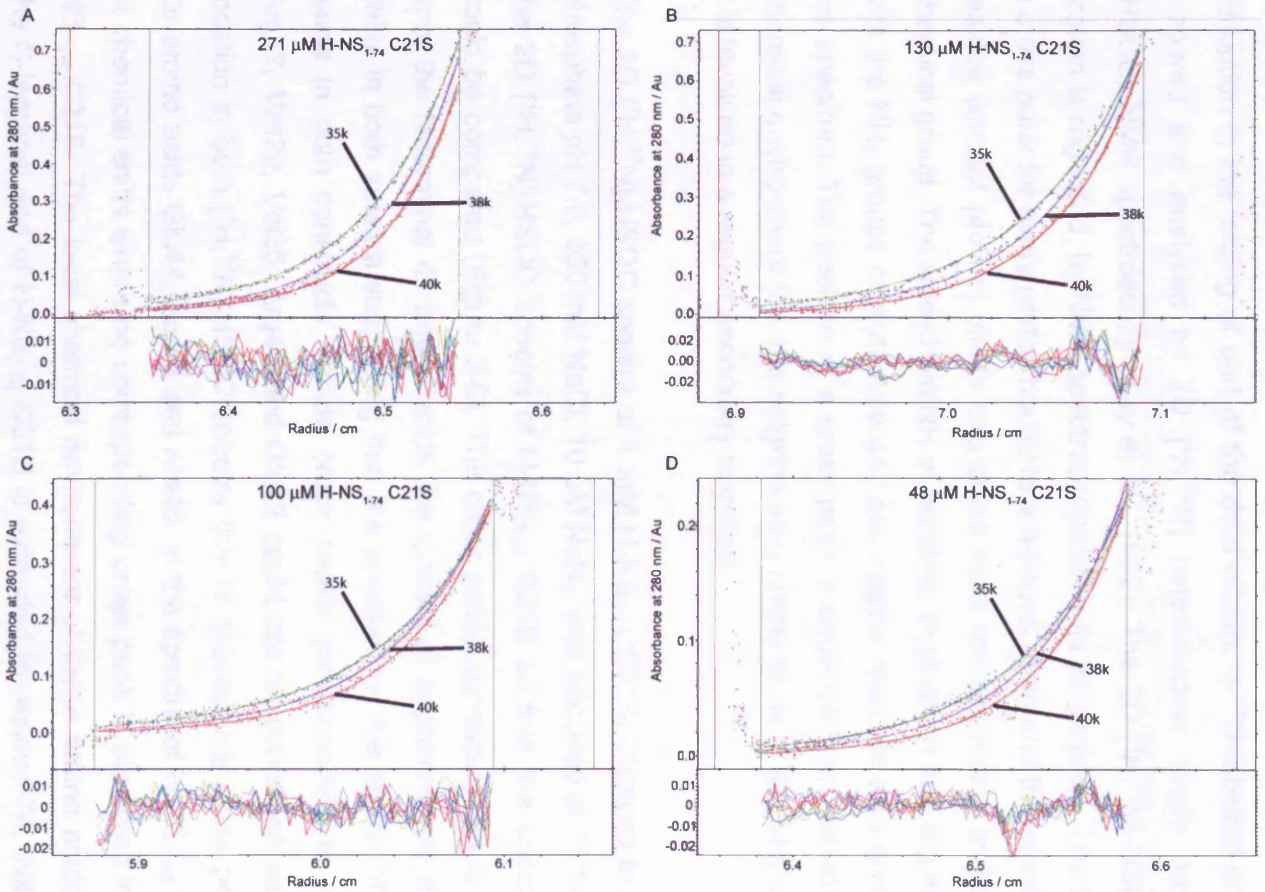


Figure 3-5. Sedimentation equilibrium analytical ultracentrifugation analysis of H-NS_{1-74} C21S in 20 mM potassium phosphate pH 7.0, 300 mM NaCl, 1 mM EDTA. Data is fitted for three different rotor speeds, 35000 rpm, 38000 rpm and 40000 rpm recorded for four protein loading concentrations 271 µM, 130 µM, 100 µM and 48 µM in a global fit set to a single species model with a molar mass of 17459.6 g mol^{-1}. Data points where the absorbance recorded is in excess of 2 absorbance units were not included in the fitting. The residuals for the fits are shown below each data set.

3.7 Assessing the Structured and Unstructured Regions of H-NS$_{1-74}$ C21S

The secondary structure predictions for H-NS form a consensus that the long helix that forms the coiled-coil dimerisation interactions spans from Leu23 to Ala67 (Figure 1-6). To test if the truncation of H-NS to produce the H-NS$_{1-74}$ C21S construct causes disruption in the folding of part of the third α-helix, a ^{15}N-labelled protein sample was prepared and analysed by 2D [^{1}H,^{15}N] heteronuclear single quantum correlation (HSQC) NMR spectroscopy (Kay et al., 1992). The 2D [^{1}H,^{15}N]-HSQC spectrum of a protein is regarded, by NMR spectroscopists as its 'fingerprint'. The spectrum contains a cross peak for every proton bound to a nitrogen atom and therefore, each amino acid residue (except proline) yields one cross peak corresponding to its backbone amide chemical group. The side-chain NH of histidine, tryptophan and arginine residues along with the NH_2 groups of glutamine and asparagine residues also give rise to signals in the spectrum. The position of a cross peak is dependent on the amino acid type, the chemical environment (*i.e.* the neighbouring amino acids in the sequence) and whether it is involved in a regular secondary structure.

The 2D [^{1}H,^{15}N]-HSQC spectra of 1 mM H-NS$_{1-74}$ C21S, disolved in 20 mM potassium phosphate pH 7.0, 300 mM NaCl, 10 μM NaN_3, was acquired at 25°C and overlaid with the 2D [^{1}H,^{15}N]-HSQC spectra of H-NS$_{2-58}$ C21S so that the spectral characteristics could be compared (Figure 3-6). The cross peaks corresponding to backbone amides from the N-terminal 41 amino acids are located at approximately the same chemical shifts in both spectra suggesting that the structure of this region of the protein is the same in both constructs. Amide cross peaks corresponding to amino acids Ala4, Arg12, Met29, Val36, Arg40 and Glu42 could not be confirmed as having the same location in both [^{1}H,^{15}N]-HSQC spectra due to overlapping cross peaks. Cross peaks for amino acids Glu44, Ser45 and Ala46, in the spectra of H-NS$_{2-58}$ C21S, are located at chemical shifts where no corresponding cross peak is observed in the spectra of H-NS$_{1-74}$ C21S. The local chemical environment of these amino acids is different when the C-terminal end of H-NS$_{2-58}$ C21S is extended to residue 74. With the exception of Ala47, Ala48, Ala49, Glu50, Val51 and Arg54, where spectral overlap precludes an unambiguous assignment of the cross peaks, residues Glu44 to Leu58 are not located in the same positions in the [^{1}H,^{15}N]-HSQC spectra of H-NS$_{2-58}$ C21S and H-NS$_{1-74}$ C21S.

Figure 3-6. Overlaid 2D [^{1}H,^{15}N]-HSQC spectra for H-NS_{2-58} C21S (red) and H-NS_{1-74} C21S (black). Amino acid and side chain assignments for H-NS_{2-58} C21S are shown as determined by (Esposito et al., 2002) The cross peak assignments corresponding to backbone or sidechain amide or amine chemical groups that are in the same position for the two protein constructs are shown in black. The assignments (for H-NS_{2-58}) shown in red indicate cross peaks for the backbone amides of amino acids that located at different chemical shifts in the two spectra or cannot be unambiguously assigned in the spectrum of H-NS_{1-74} C21S due to overlap of cross peaks.

The structure of the C-terminal end of H-NS_{2-58} C21S is altered, by extending the protein to residue 74. The predicted helix spanning from Leu23 to Ala67 (Figure 1-6) that forms the coiled-coil core of the oligomerisation domain is truncated in the H-NS_{2-58} construct. The [^{1}H,^{15}N]-HSQC for H-NS_{1-74} C21S shows considerable cross peak overlap in the spectra, with many of the backbone amide cross peaks situated within the range of 7.7 ppm to 8.3 ppm in the proton dimension. A concentration of cross peaks in this region can be caused by either an unstructured region, a predominantly α-helical secondary structure or an absence of aromatic residues in the protein. In the 2D [^{1}H,^{15}N]-HSQC spectra of H-NS_{1-74} C21S the majority of the cross peaks have very similar intensity suggesting that the majority of the protein is structured but seven intense cross peaks are observed in the HSQC that could indicate a disordered region.

To investigate whether there are amino acids located in unstructured regions in the H-NS_{1-74} C21S protein, a (^{1}H)-^{15}N heteronuclear NOE experiment was performed on the same protein NMR sample. This was performed to determine if the overlapping cross peaks in the [^{1}H,^{15}N]-HSQC spectrum of H-NS_{1-74} C21S, found between 7.6 ppm and 8.3 ppm for the proton frequency, The steady state ^{15}N NOE experiment provides a measure of the rate of amide bond vector rotation (Farrow et al., 1994) and thus can be used to identify amino acids that are unstructured or highly mobile within a protein. The steady state (^{1}H)-^{15}N heteronuclear NOE (hetNOE) is dependent upon a mechanism of relaxation termed cross-relaxation. The purpose of this experiment is to observe the maximum "enhancement" of magnetisation of the ^{15}N nucleus as the ^{1}H spin is saturated. The hetNOE is measured with two separate 2D heteronuclear correlation experiments. One experiment includes a saturation period of the ^{1}H spins; the other experiment omits this period. The intensity of the ^{15}N magnetisation can then be compared in the two spectra to determine the relative enhancement due to ^{1}H saturation. Mobile amides or amines give rise to a significantly diminished or negative cross peaks upon ^{1}H spin saturation.

Figure 3-7 shows the two hetNOE spectra, (a) without and (b) with ^{1}H spin saturation, for H-NS_{1-74} C21S. The (^{1}H)-^{15}N heteronuclear NOE data were recorded with 3.0 seconds ^{1}H saturation in the latter part of a 3.5 seconds preparation period delay, which was also used without radio frequency pulses for the reference 2D spectrum without NOE. The sidechain NH cross peaks, indicated by the horizontal black lines, are significantly reduced in intensity due to their increased mobility relative to the rest of the protein. In addition there are eight backbone amide cross peaks that have negative intensity, coloured red, (peaks labelled 3, 5, 6, 7 and 8) or are not visible (peaks labelled 2, 4 and 9) in the hetNOE spectrum with ^{1}H spin saturation (Figure 3-

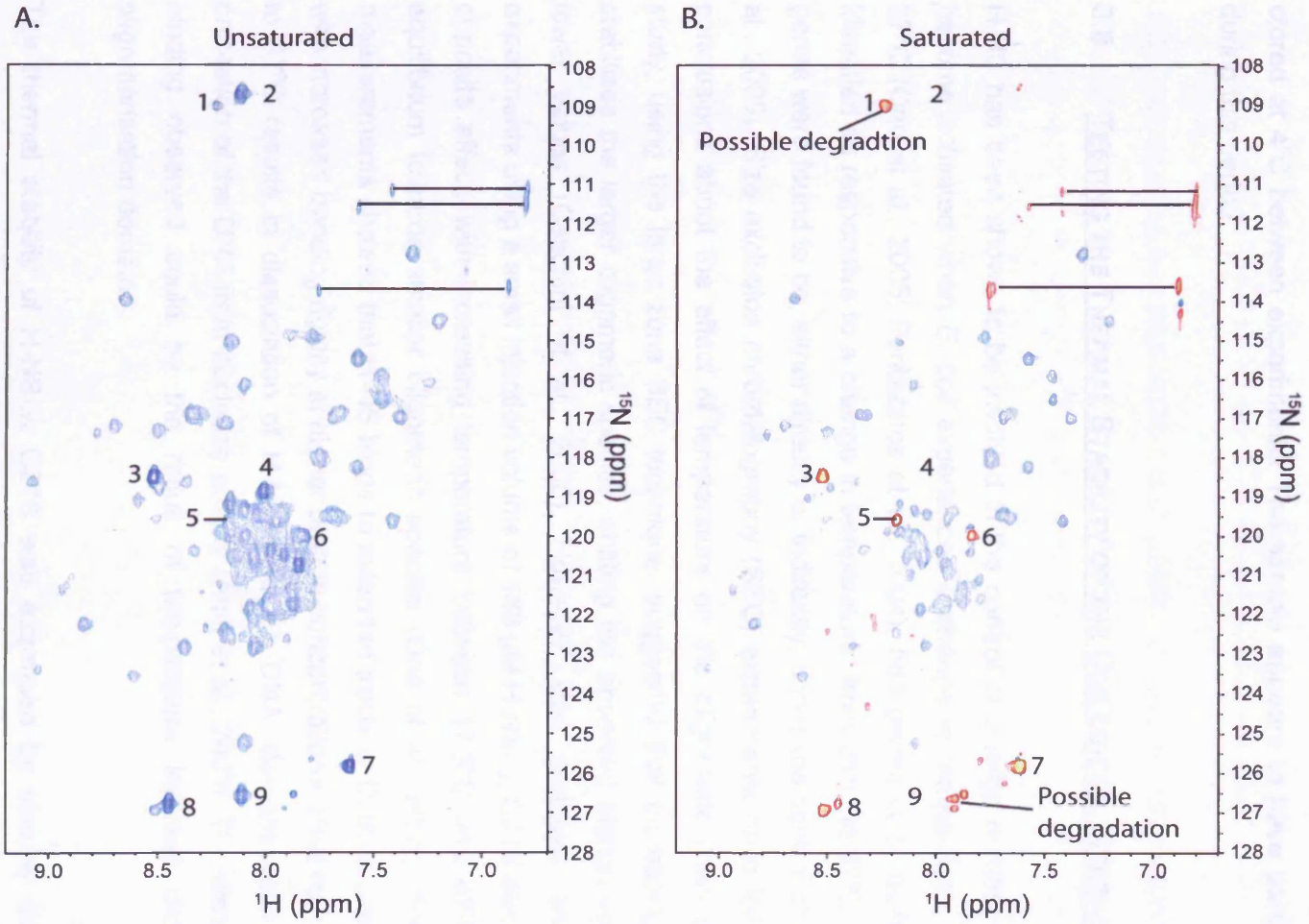


Figure 3-7. 2D (^{1}H)-^{15}N heteronuclear overhauser effect spectra for H-NS$_{1-74}$ C21S. Positive intensity cross peaks are coloured blue with negative intensity cross peaks coloured red. The numbers 1 to 9 indicate nine cross peaks in the spectra that have significantly reduced intensity in the ^{1}H spin saturated hetNOE spectrum relative to the unsaturated spectrum.

7b). All of these cross peaks were unassigned, through comparison to the [^{1}H,^{15}N]-HSQC spectrum of H-NS$_{2-58}$ C21S, and, therefore, are probably from amides of unstructured residues at the C-terminus of H-NS$_{1-74}$ C21S. The cross peak labelled 1 is not present in the [^{1}H,^{15}N]-HSQC spectra of H-NS$_{1-74}$ C21S and is therefore likely to be the result of degradation of the C-terminal end of the protein. The hetNOE experiment was performed a week after the [^{1}H,^{15}N]-HSQC experiment with the protein sample stored at 4°C between experiments. The sample appears to have partially degraded during this period.

3.8 TESTING THE THERMAL STABILITY OF THE COILED-COIL INTERACTION

H-NS has been shown to be involved in the control of a large number of genes that become activated when *E. coli* experiences a change in temperature, from 25°C to 37°C (Ono et al., 2005; Pantazatos et al., 2004). 581 genes in *S. typhimurium* were identified as responsive to a change in temperature, from 25°C to 37°C; 77% of these genes were found to be, either directly or indirectly, under the control of H-NS (Ono et al., 2005). Size exclusion chromatography (SEC) experiments have led to conflicting conclusions about the effect of temperature on the oligomeric state of H-NS. One study, using the large zone SEC technique, suggested that increasing temperature stabilises the larger oligomeric species, shifting the observed elution volume towards lower values (Ceschini et al., 2000). However, size exclusion chromatography experiments using a small injection volume of 168 µM H-NS$_{1-90}$ C21S demonstrated the opposite affect, with increasing temperature between 17.5°C and 45°C shifting the equilibrium towards smaller oligomeric species (Ono et al., 2005). Force extension measurements showed that H-NS binds to extended tracts of DNA to create a complex with increased bending rigidity at higher protein concentrations. Heating of the complex to 37°C results in dissociation of H-NS from the DNA, demonstrated by a loss of protection of the DNA from nuclease activity (Amit et al., 2003). The disruption of DNA binding observed could be the result of temperature induced changes in the oligomerisation domain.

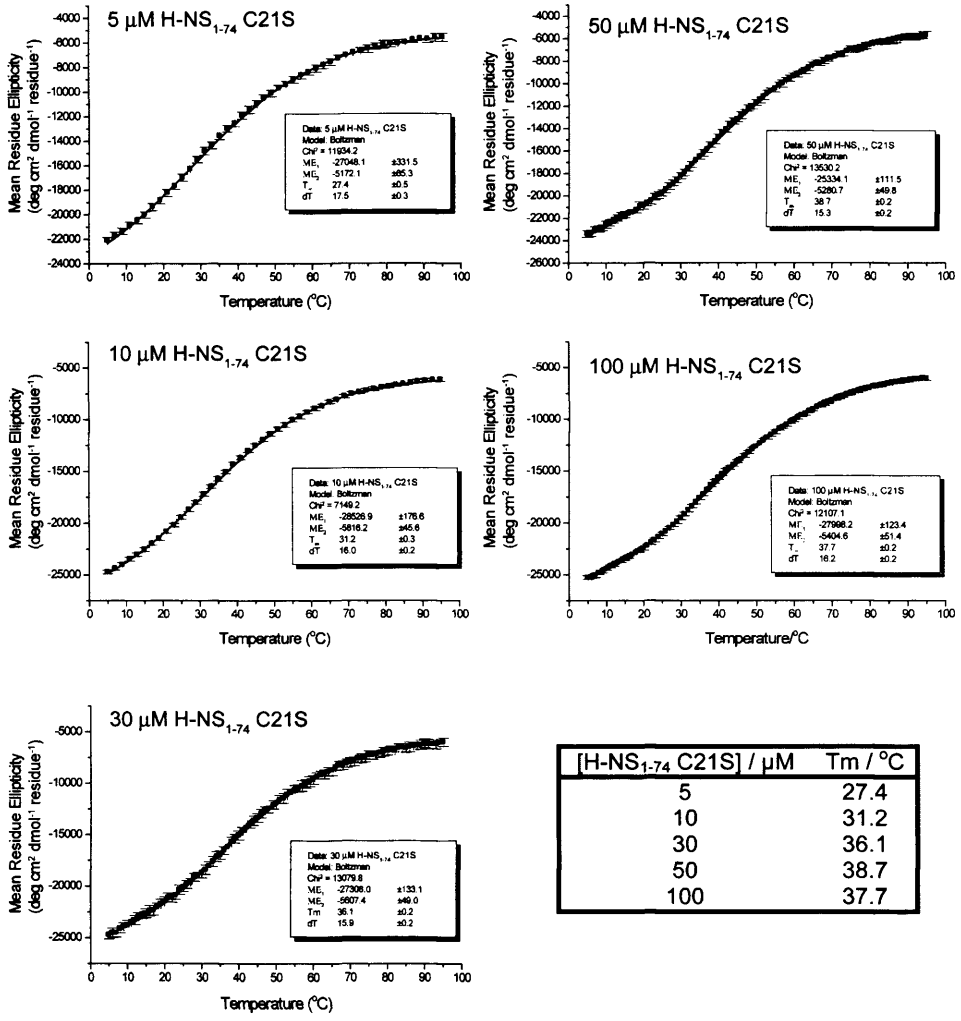
The thermal stability of H-NS$_{1-74}$ C21S was examined by circular dichroism. This protein construct contains the complete predicted coiled-coil region of H-NS but is disrupted in its ability to form higher order oligomeric species, making the construct suitable for testing the thermal stability of the coiled-coil interaction. Various concentrations ranging from 5 µM to 100 µM of H-NS$_{1-74}$ C21S were prepared in 20 mM potassium phosphate pH 7.0, 100 µM NaN_3. The melting of the protein samples

was monitored at 222 nm using a circular dichroism spectropolarimeter (Figure 3.8). The melting point was determined by fitting the raw data to a Boltzman distribution equation

$$MRE_{222} = \frac{MRE_1 - MRE_2}{1 + e^{\frac{T - T_m}{dT}}} + MRE_2$$

MRE_{222} is the mean residue ellipticity of at 222 nm, MRE_1 and MRE_2 are the lower and upper asymptotes of the fit respectively, T is the temperature in degress Celcius, T_m is the melting temperature and dT is the slope at the mid point of the transition (at the T_m). The melting temperature for this protein is concentration dependent in the concentration range of 5 μM to 30 μM, changing from 27.4°C to 36.1°C respectively, which is in good agreement with the concentration dependence of the far UV circular dichroism spectra observed for H-$NS_{1\text{-}65}$ C21S and H-NS_{FL} C21S proteins (Smyth et al., 2000). At concentrations above 30 μM the melting temperature of the construct remains constant at approximately 37°C. The AUC results, described in section 3.6, demonstrate that H-$NS_{1\text{-}74}$ C21S forms a stable homodimer, at concentrations between 48 μM and 271 μM. The concentration dependence of the melting temperature, of H-$NS_{1\text{-}74}$ C21S, at concentrations below 30 μM, suggests that between 5 μM and 30 μM this protein is in equilibrium between monomer and homodimer.

The TlpA protein from *S. typhimurium* has a temperature sensitive coiled-coil that undergoes a fully reversible transition from a folded coiled-coil to a unfolded monomer when the temperature is changed from 25°C to 55°C (Naik *et al.*, 2001). It has been suggested that this protein's coiled-coil acts as a temperature sensor to modulate the transcription repression of its target genes in a temperature sensitive manner (Hurme *et al.*, 1997). The melting of TlpA is cooperative, with a rapid transition between the folded protein and unfolded monomer between 35°C and 45°C. The melting of H-$NS_{1\text{-}74}$ C21S shows a gradual transition between the folded homodimer and the unfolded monomer suggesting that the unfolding of the protein is non-cooperative.



[H-NS1-74 C21S] / μM	Tm / °C
5	27.4
10	31.2
30	36.1
50	38.7
100	37.7

Figure 3-8. Circular dichroism melt curves for H-NS_{1-74} C21S at various protein concentrations from 5 μM to 100 μM measured at 222 nm in 20 mM potassium phosphate pH 7.0, 100 μM NaN_3. The error bars show the range of values observed during the 10 second sampling time for each data point. The melting temperature for each concentration of H-NS_{1-74} C21S is shown in the table. The red line indicates the non-linear least squares fit for each dataset.

3.9 $StpA_{1-65}$ FORMS A DISCRETE HOMODIMER

StpA is a paralogue of H-NS with 57% sequence identity with H-NS. With such close sequence homology it seemed likely that StpA constructs that are truncated to remove the C-terminal end of the oligomerisation domain would form discrete dimers as had been previously shown for H-NS_{1-65} (Smyth et al., 2000; Badaut et al., 2002). To test this hypothesis, various concentrations of $StpA_{1-65}$ were analysed by sedimentation equilibrium AUC and the apparent molecular mass of protein was determined. The results for four protein concentrations between 71.2 μM and 569 μM at three separate rotor speeds, 35000 rpm, 38000 rpm and 40000 rpm were fitted, using a global fit, to a single species model with molar mass of 16476.98 g mol^{-1} (Figure 3-9). This is the molar mass of a $StpA_{1-65}$ homodimer.

3.10 ANALYSIS OF THE HIGH ORDER OLIGOMERISATION POTENTIAL OF H-NS_{1-83} C21S

The longer H-NS_{1-77} C21S and H-NS_{1-83} C21S constructs are both able to form species larger than homodimers based on the size exclusion chromatography results discussed in section 3.3.4. Whilst the self association of wild type H-NS to form homodimers has been shown previously, by cross-linking and AUC experiments (Badaut et al., 2002; Ueguchi et al., 1996), there remains significant debate within the literature with regard to the size and distribution of the higher order oligomeric species. Large zone SEC experiments showed that H-NS was in equilibrium between monomer, dimer and tetramer species with only a small population of homodimer present under equilibrium conditions (Ceschini et al., 2000). However other SEC studies suggested that wild type H-NS forms complexes as large as 20-mers in a concentration dependent manner (Smyth et al., 2000). In both experiments, proteins whose shape approximates to a sphere were used to calibrate the column. However the determination of the molar mass of H-NS by size exclusion chromatography is unreliable because H-NS has an elongated structure (Badaut et al., 2002) and the higher order oligomeric species readily dissociate as the protein is diluted in the SEC column (Smyth et al., 2000).

To investigate the oligomeric species formed by the H-NS oligomerisation domain the H-NS_{1-83} C21S construct was used because it is more resistant to proteolysis than the shorter H-NS_{1-77} C21S construct and contains all of the predicted residues for the fourth α-helix (Pro72 to Ala81) of H-NS. In addition, samples of this protein in aqueous buffers

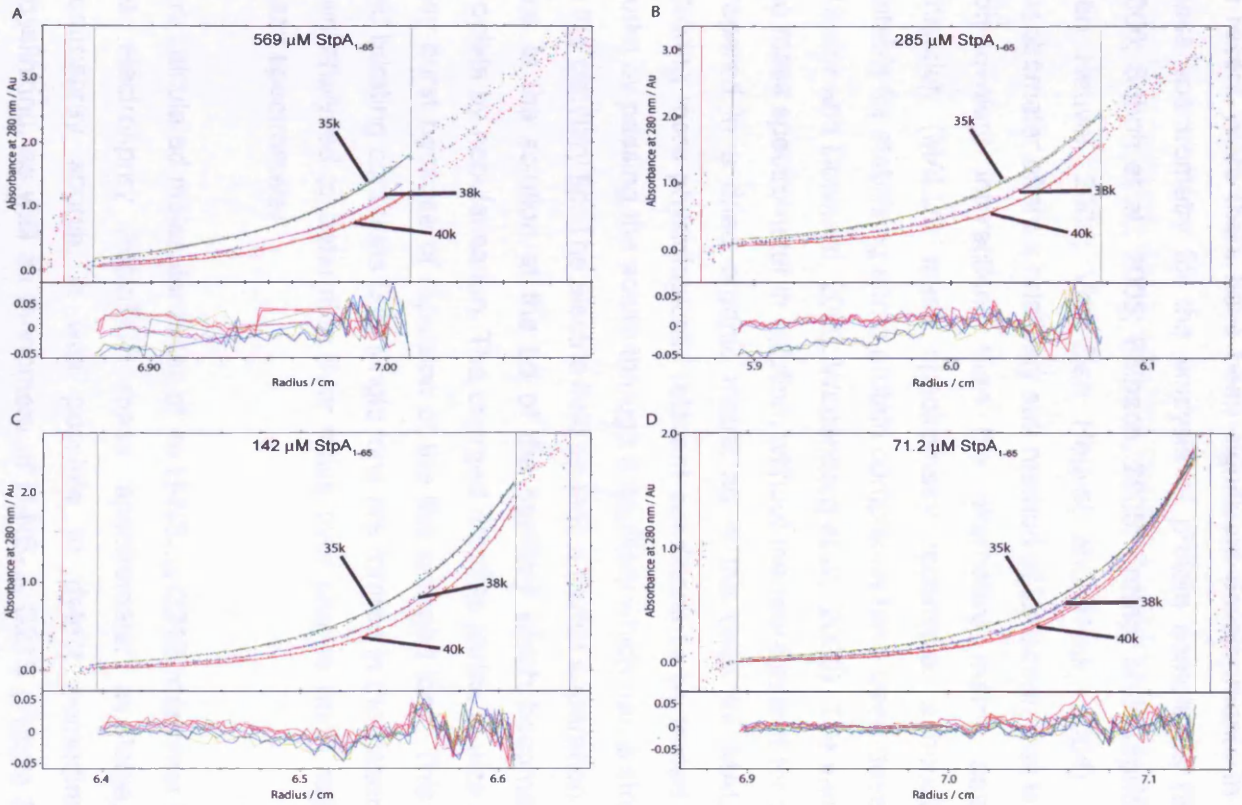


Figure 3-9. Sedimentation equilibrium analytical ultracentrifugation analysis of $StpA_{1-65}$ in 20 mM potassium phosphate pH 7.0, 300 mM NaCl, 1 mM EDTA. Data is fitted for three different rotor speeds, 35000 rpm, 38000 rpm and 40000 rpm recorded for four protein concentrations 569 µM, 285 µM, 142 µM and 71.2 µM in a global fit set to a single species model with a molar mass of 16476.98 g mol^{-1}. Data points where the absorbance recorded is in excess of 2 absorbance units were not included in the fitting. The residuals for the fits are shown below each data set.

can be concentrated up to 2 mM, without signs of precipitation, allowing a wide range of concentrations to be tested.

3.10.1 Mass Spectrometry of H-NS_{1-83} C21S under Non-Covalent Interaction Conditions.

In recent years there have been significant developments in the use of electrospray mass spectrometry for the analysis of protein complexes (Benesch and Robinson, 2006; Sobott et al., 2005; Bolbach, 2005; Farmer and Caprioli, 1998; Heck and Van Den Heuvel, 2004; Van Den Heuvel and Heck, 2004). An electrospray mass spectrometer offers a relatively soft method of ionisation that is less prone to disrupting non-covalent interactions than the alternative matrix assisted laser desorption ionisation (MALDI) mass spectrometry technique, although ionisation conditions suitable for stabilising some protein complexes have been developed for this technique (Kiselar and Downard, 2000; Wattenberg et al., 2000). The sample is also injected into the mass spectrometer in solution, without the requirement for the protein sample to be prepared in a dried organic matrix as is the case for MALDI mass spectrometry, allowing more physiologically relevant conditions to be tested. Electrospray ionisation works by passing the solute through a capillary which has a strong electric field applied to the capillary tip. The electric field causes a partial separation of positive and negative ions in the solution at the tip of the capillary which becomes released as charged droplets by nebularisation. The charged droplets shrink in size due to evaporation until they burst because of repulsion of the like charged ions. This process of evaporation and bursting continues until single ions are formed in the gaseous phase. The ions are then analysed to determine their mass over charge (m/z) ratio using a time of flight mass spectrometer.

The calculated molecular mass of an H-NS_{1-83} C21S monomer is 9783.1 g mol^{-1}. Using the electrospray microTOF mass spectrometer available, with an orthoganol electrospray source, it was possible to detect homodimers, without chemical crosslinking, as well as monomers of H-NS_{1-83} C21S (Figure 3-10) but larger species were not identified. The mass spectrometer settings used for the experiments are described in section 2.5.2. The ionisation source used is not optimal for the detection of non-covalently bound protein complexes. Ideally a nanoflow source (50 nL min^{-1}) would be used, to reduce the level of solvent that needs to be evaporated rather than the electrospray source set at 100 µL min^{-1} used in these experiments. The evaporation of solvent is a significant problem when using aqueous buffers; the rate of

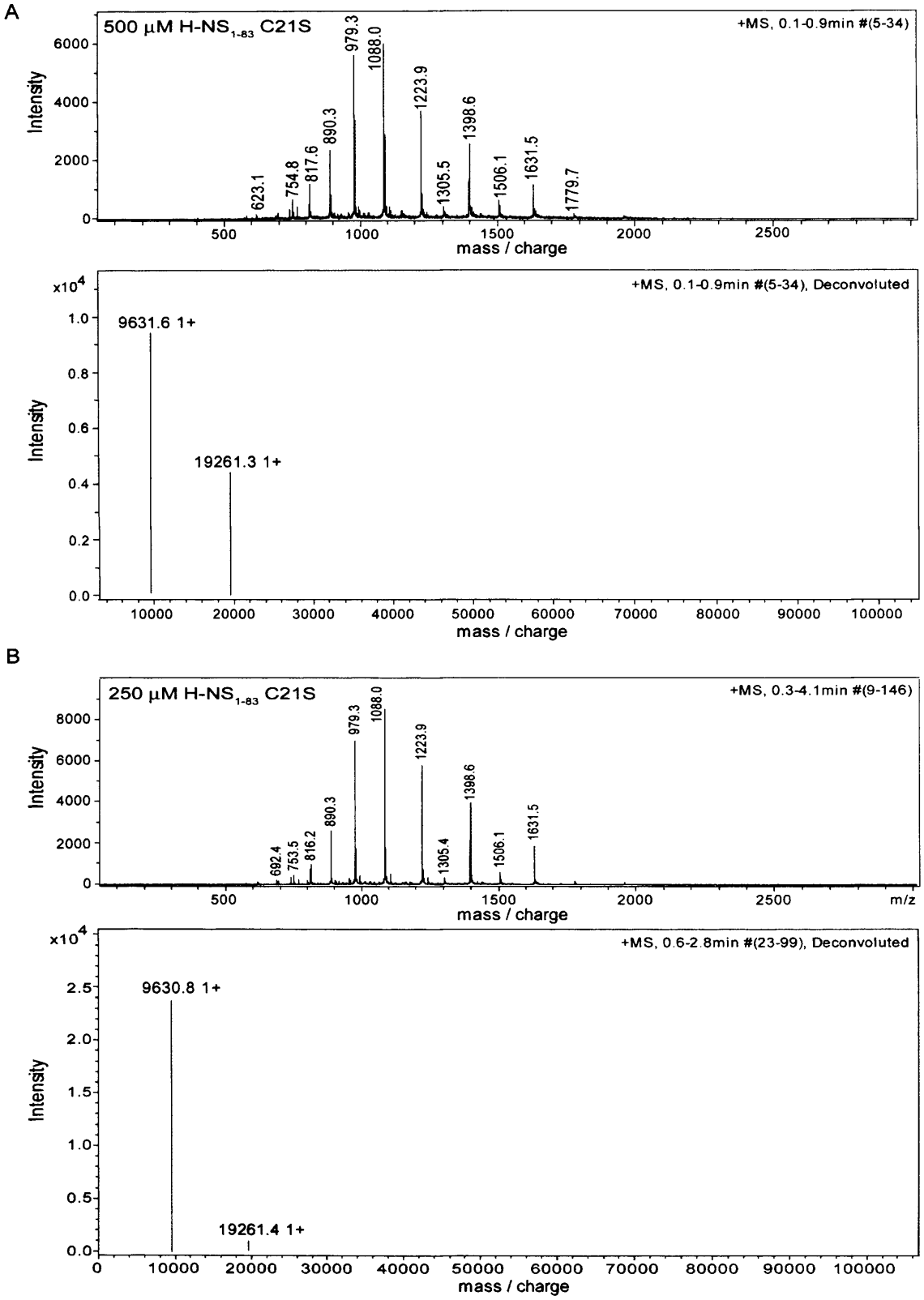


Figure 3-10. Mass Spectrometry of H-NS_{1-83} C21S at 500 µM (A) and 250 µM (B) dissolved in 200 mM ammonium acetate pH 7.0. The raw data collected by the mass spectrometer after manual injection of the sample is shown above the deconvoluted mass spectra.

evaporation is slow compared to the more volatile organic solvents used in denaturing conditions. With less evaporation the amount of ions entering the mass spectrometer is reduced resulting in lower signal to noise levels. Additionally, the temperature of the capillary was reduced to 37°C, the lowest temperature available, in an attempt to stabilise higher order oligomeric species of H-NS_{1-83} C21S but this results in a further loss of signal due to reduced evaporation of the solvent.

3.11 Changing the Ionic Strength of the Solvent Affects the Oligomeric State of H-NS_{1-83} C21S

The oligomeric state of 200 µM H-NS_{1-83} C21S, in 20 mM MOPS pH 7.0 supplemented with 10 mM, 50 mM, 100 mM, 150 mM and 200 mM $NaClO_4$ was examined by sedimentation velocity AUC. All of the samples were tested at 20°C using a rotor speed of 40000 rpm. $NaClO_4$ was used to increase the ionic strength of the solvent because the concentration of perchlorate ions in solution can be accurately determined. Chloride ions were not used because the form polyhalocations in solution (such as HCl_2^-) making the ionic strength more difficult to calculate.

The sedimentation boundary profiles for each protein sample were analysed, using the c(s) distribution analysis in SEDPHAT (Schuck, 2003), to assess the effect of different concentrations of $NaClO_4$ on the distribution of H-NS_{1-83} C21S species in the protein sample. A global fit was used for all six $NaClO_4$ concentrations allowing the weight average frictional ratio to be optimised by non-linear regression. The raw data and the fitted curves are shown in Figure 3-11. The best fit for the raw data was achieved with a weight average frictional ratio of 2.16. Although the presence of multiple species in solution means that this frictional ratio provides only an estimate of the frictional ratio of the different oligomeric species present, it does indicate the presence of elongated proteins in the solution. The c(s) distribution for each concentration of $NaClO_4$ tested (Figure 3-12) shows a shift towards larger oligomeric species, indicated by a shift in the distribution towards higher sedimentation coefficient species, when the concentration of $NaClO_4$ in the solvent is increased from 10 mM to 100 mM. As the $NaClO_4$ concentration is increased in the solvent the density and viscosity of the solvent also increases so protein species will sediment more slowly as the $NaClO_4$ concentration increases. The sedimentation coefficient of an H-NS_{1-83} C21S homodimer, for example, will decrease as the concentration of $NaClO_4$ increases. To correct for changes in the solvent, the peaks in the c(s) distribution were integrated to obtain the weight average

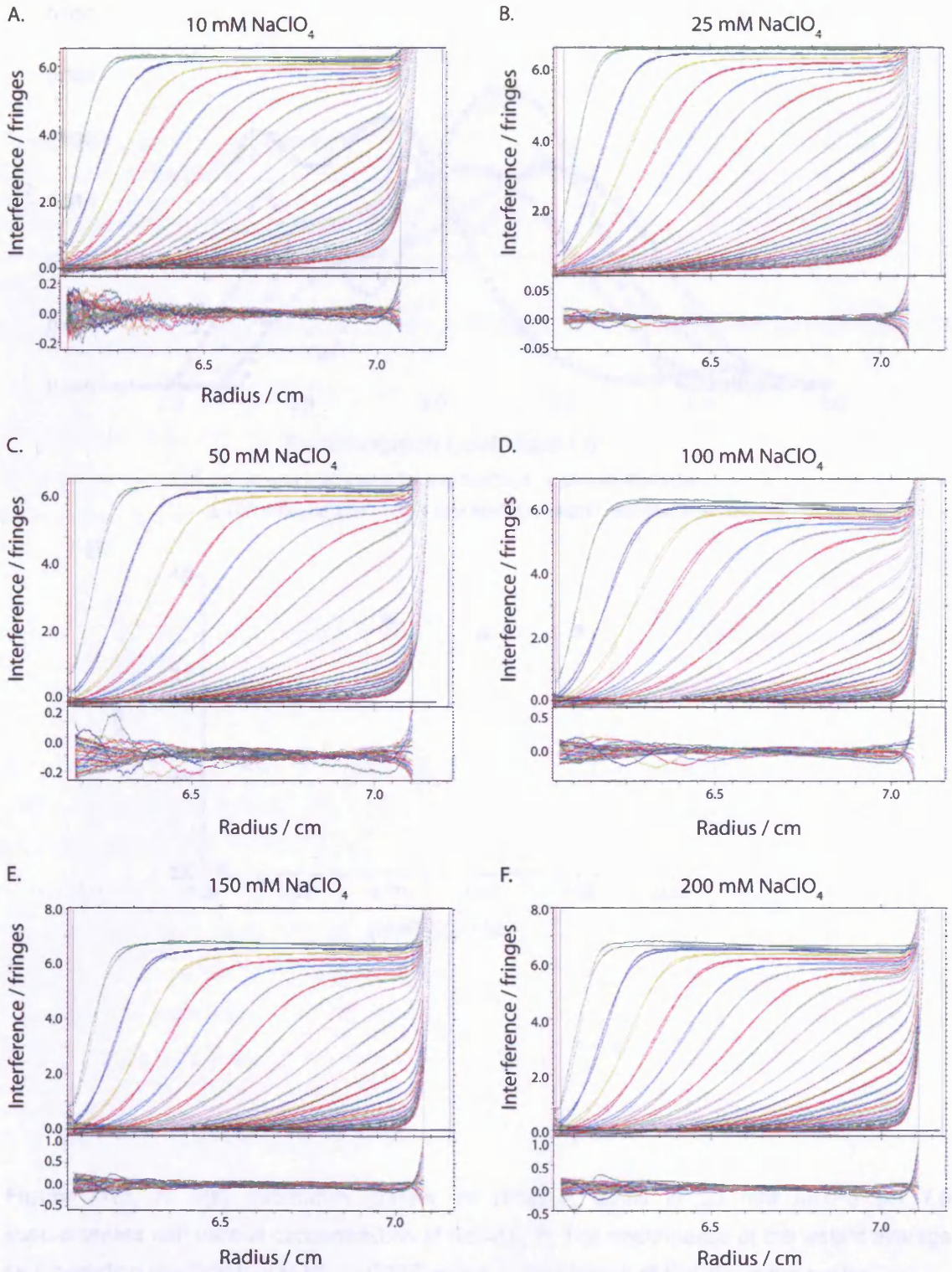


Figure 3-11. Sedimentation velocity analysis of 200 μM H-NS_{1-83} C21S in 20 mM MOPS pH 7.0 supplemented with 10 mM (A), 25 mM (B), 50 mM (C), 100 mM (D) 150 mM (E) and 200 mM (F) $NaClO_4$. Each dataset shows the raw sedimentation scans at 40000 rpm fitted using the hybrid local continuous model implemented using SEDPHAT (Schuck, 2000). For clarity, only every fifth scan is shown. The residuals for the curve fits are shown below each dataset. The c(s) distribution graphs generated from this analysis are shown in Figure 3-12.

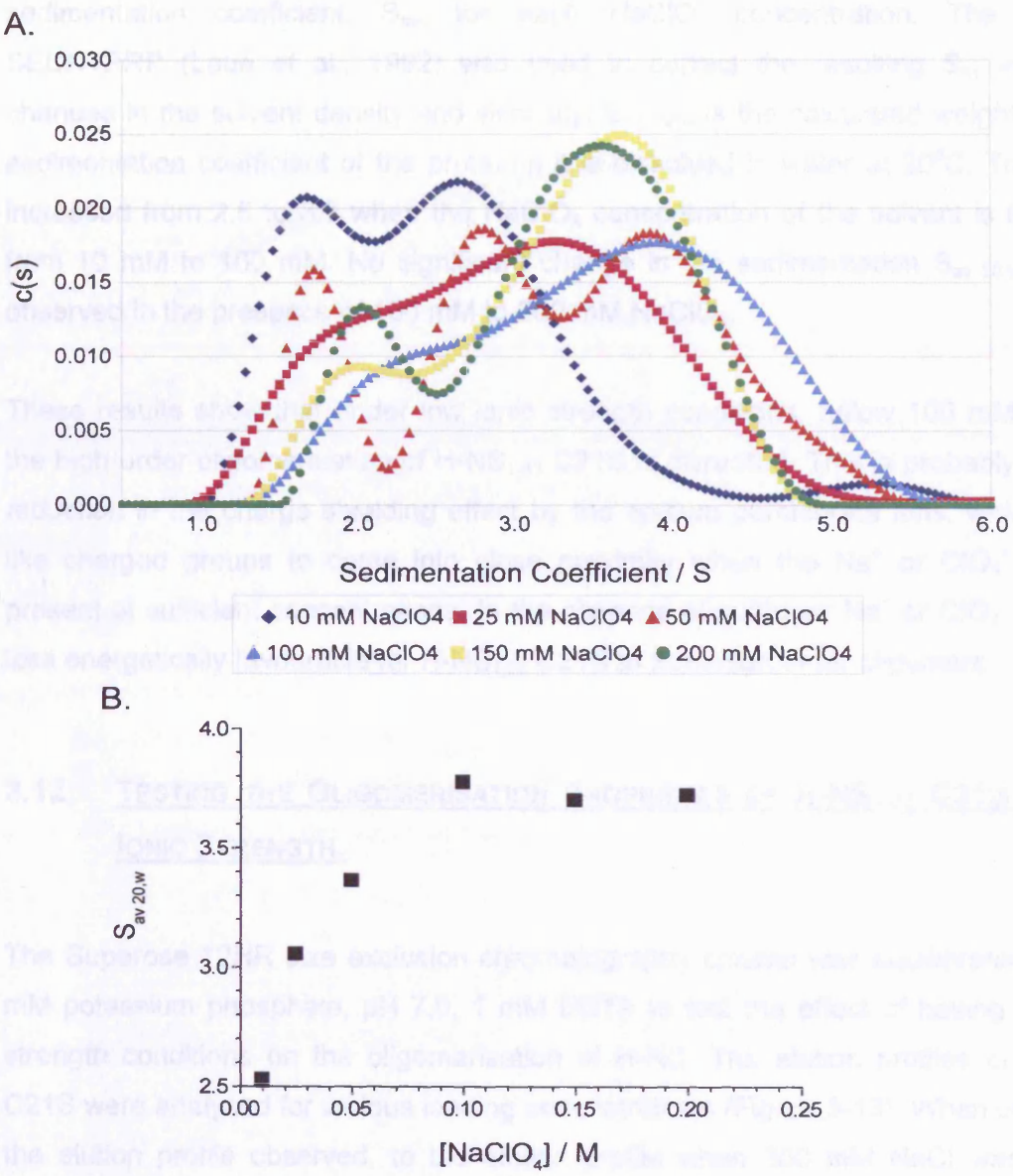


Figure 3-12. A. c(s) distribution graphs for H-NS_{1-83} C21S in 20 mM MOPS pH 7.0 supplemented with various concentrations of $NaClO_4$. B. The dependence of the weight average sedimentation coefficient of H-NS_{1-83} C21S on the concentration of $NaClO_4$ in the solvent.

sedimentation coefficient, S_{av}, for each $NaClO_4$ concentration. The program SEDNTERP (Laue et al., 1992) was used to correct the resulting S_{av} values for changes in the solvent density and viscosity. $S_{av\ 20,w}$ is the calculated weight average sedimentation coefficient of the protein if it is dissolved in water at 20°C. The $S_{av\ 20,w}$ increased from 2.5 to 3.7 when the $NaClO_4$ concentration of the solvent is increased from 10 mM to 100 mM. No significant change in the sedimentation $S_{av\ 20,w}$ value is observed in the presence of 100 mM to 200 mM $NaClO_4$.

These results show that under low ionic strength conditions, below 100 mM $NaClO_4$, the high order oligomerisation of H-NS_{1-83} C21S is disrupted. This is probably due to a reduction in the charge shielding effect by the sodium perchlorate ions, which allows like charged groups to come into close proximity when the Na^+ or ClO_4^- ions are present at sufficient concentrations. In the absence of sufficient Na^+ or ClO_4^- ions, it is less energetically favourable for H-NS_{1-83} C21S to form high order oligomers.

3.12 Testing the Oligomerisation Properties of H-NS_{1-83} C21S at Low Ionic Strength.

The Superose 12HR size exclusion chromatography column was equilibrated with 20 mM potassium phosphate, pH 7.0, 1 mM EDTA to test the effect of having low ionic strength conditions on the oligomerisation of H-NS. The elution profiles of H-NS_{1-83} C21S were analysed for various loading concentrations (Figure 3-13). When comparing the elution profile observed, to the elution profile when 300 mM NaCl was present (Figure 3-3) there is a shift towards larger elution volumes and the peaks for the protein have a more symmetrical profile. The sample appears more homogeneous under low ionic strength conditions.

The low ionic strength conditions disrupt the high order oligomerisation interactions, shifting the equilibrium towards the formation of homodimers. A calibration curve based on the elution volume of globular protein standards for the molar mass range of 13.7 kDa to 200 kDa was prepared in the same buffer conditions to account for the change in elution volume caused by the change in solvent conditions (Figure 2-16). In the absence of NaCl in the solvent all of the globular protein standards elute at a later elution volume than the elution volume observed when 300 mM NaCl was present.

Figure 3-14 shows the concentration dependence of the apparent molecular weight of H-NS_{1-83} C21S in the presence and absence of 300 mM NaCl showing that in the

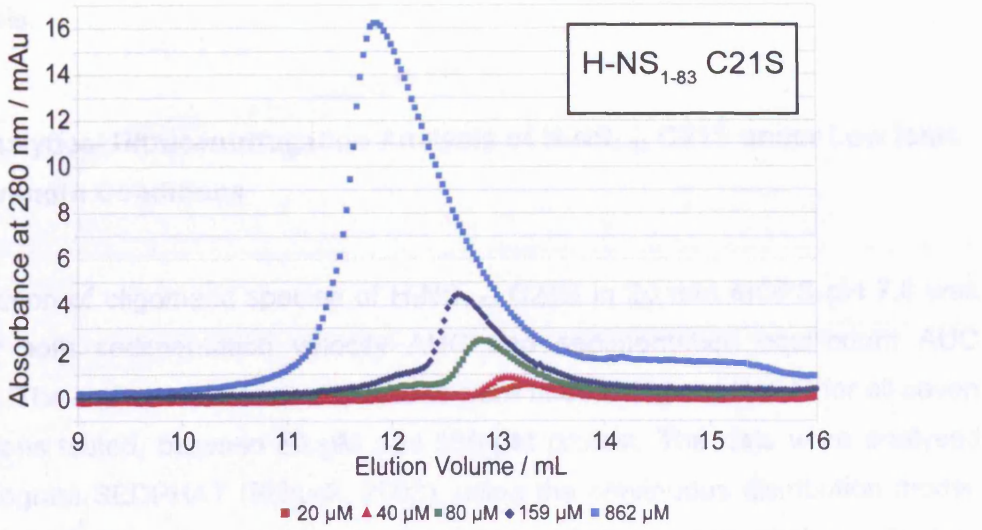


Figure 3-13. Analytical size exclusion chromatography titration of H-NS_{1-83} C21S under low ionic strength conditions (20 mM potassium phosphate, pH7.0). 100 µL protein samples were injected onto a pre-equilibrated Superose 12HR column at a flow rate of 0.75 mL min^{-1}. The elution profile was monitored spectrophotometrically at 280 nm.

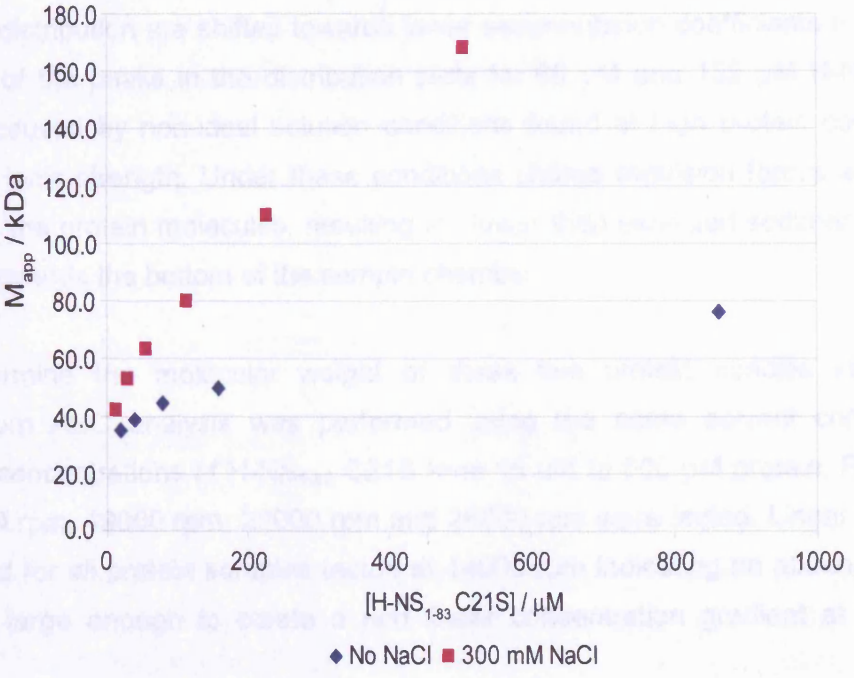


Figure 3-14. The apparent molar mass of H-NS_{1-83} C21S determined by size exclusion chromatography, using a globular protein calibration curve. The graph shows the apparent molar mass of H-NS_{1-83} C21S in 20 mM potassium phosphate pH 7.0 (blue) and 20 mM potassium phosphate pH 7.0, 300 mM NaCl (pink).

absence of any NaCl there is disruption to the formation of high order oligomeric species confirming the effect of low ionic strength observed by sedimentation velocity AUC analysis.

3.12.1 Analytical Ultracentrifugation Analysis of H-NS_{1-83} C21S under Low Ionic Strength Conditions

The distribution of oligomeric species of H-NS_{1-83} C21S in 20 mM MOPS pH 7.0 was studied by both sedimentation velocity AUC and sedimentation equilibrium AUC techniques. The sedimentation velocity results were fitted using a global fit for all seven concentrations tested, between 69 µM and 559 µM protein. The data were analysed with the program SEDPHAT (Schuck, 2003), using the continuous distribution model, c(s) (Figure 3-15). The curves were normalised, by dividing the concentration values by the total protein concentration present in the c(s) distribution. The c(s) distribution plots for the protein samples between 69 µM and 132 µM show the presence of two protein species with sedimentation coefficients of approximately 1.5 S and 2.3 S. As the protein concentration increases the proportion of protein corresponding to the 2.3 S species increases showing a concentration dependent self-association interaction. At the higher concentrations of 281 µM and 559 µM H-NS_{1-83} C21S the peaks observed in the c(s) distribution are shifted towards lower sedimentation coefficients relative to the position of the peaks in the distribution plots for 69 µM and 132 µM H-NS_{1-83} C21S. This is caused by non-ideal solution conditions found at high protein concentrations and low ionic strength. Under these conditions charge repulsion forces are observed between the protein molecules, resulting in slower than expected sedimentation of the protein towards the bottom of the sample chamber.

To determine the molecular weight of these two protein species sedimentation equilibrium AUC analysis was performed using the same solvent conditions with various concentrations of H-NS_{1-83} C21S from 16 µM to 500 µM protein. Rotor speeds of 14000 rpm, 19000 rpm, 22000 rpm and 26000 rpm were tested. Linear curves were observed for all protein samples tested at 14000 rpm indicating an absence of protein species large enough to create a non linear concentration gradient at this angular velocity.

Figure 3-16 shows the protein concentration and rotor speed dependence of the apparent molar mass of H-NS_{1-83} C21S in 20 mM MOPS pH 7.0. As the protein concentration is increased the apparent molar mass increases from 17.2 kDa to 35.1 kDa for the range of concentrations shown. The apparent molar mass (M_{app}) was

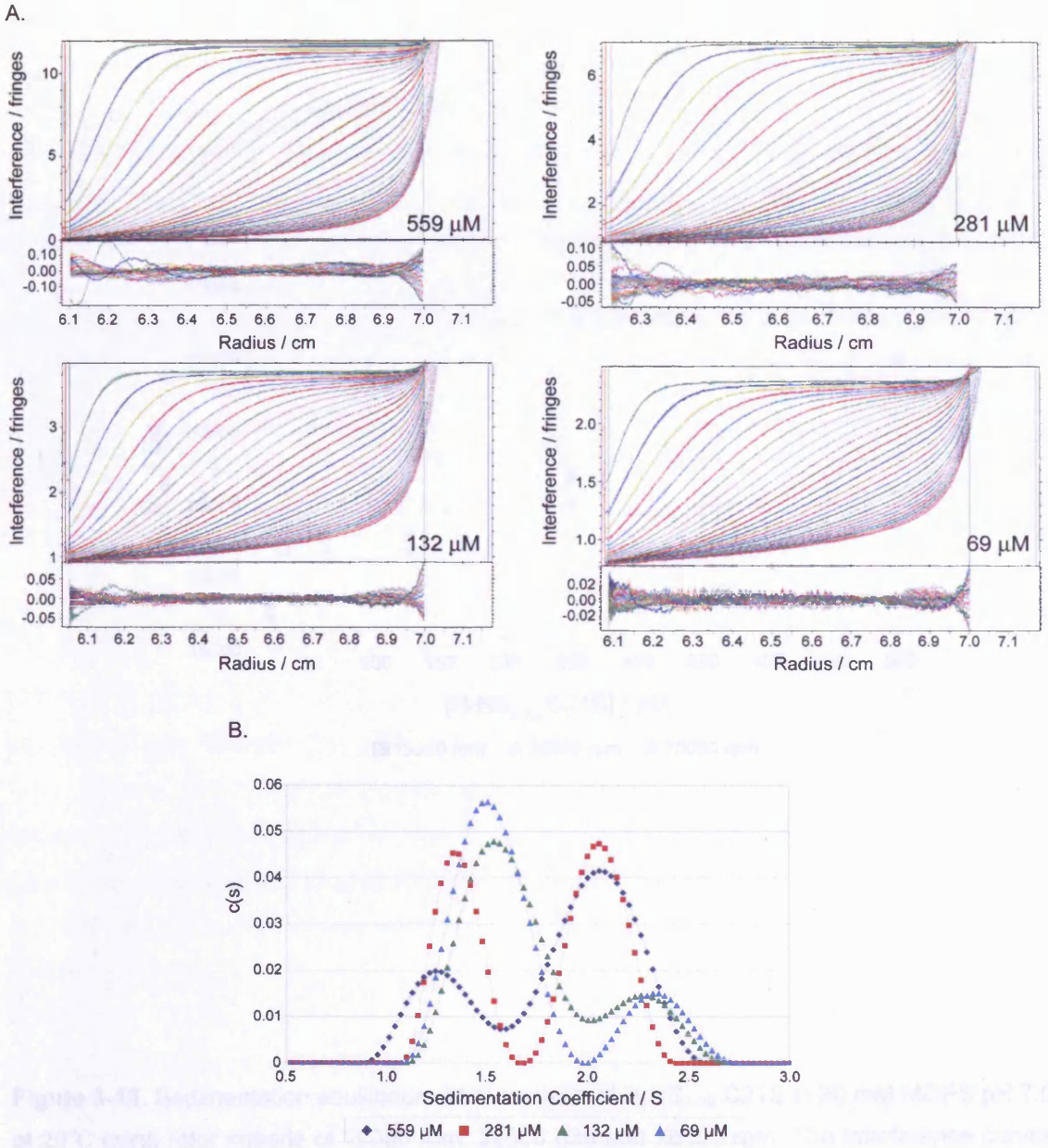


Figure 3-15. Sedimentation velocity analytical ultracentrifugation analysis of H-NS_{1-83} C21S in 20 mM MOPS pH 7.0. A. The raw data for 559 µM, 281 µM, 132 µM and 69 µM H-NS_{1-83} C21S protein samples is shown in the upper pane for each protein concentration. The residuals for the fitted curves are shown below each dataset. The sedimentation boundaries were fitted using the c(s) distribution analysis implemented in the SEDPHAT software (Schuck, 2003). For clarity only the sedimentation boundaries for every third time point are shown. The vertical red and blue lines indicate the positions of the meniscus and the bottom of the sample cell respectively. The two vertical green lines indicate the data range used in the analysis. B. The normalised c(s) distribution.

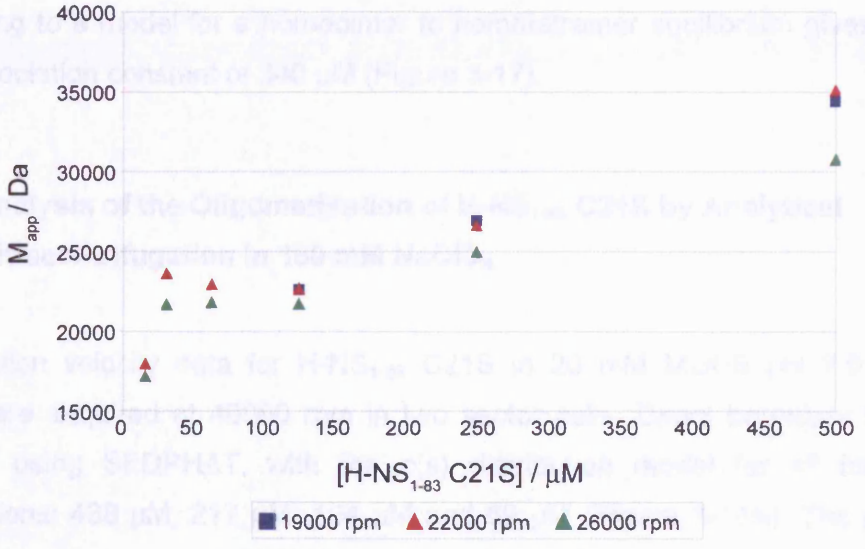


Figure 3-16. Sedimentation equilibrium AUC analysis of H-NS$_{1-83}$ C21S in 20 mM MOPS pH 7.0 at 20°C using rotor speeds of 19000 rpm, 22000 rpm and 26000 rpm. The interference curves showing the distribution of the protein at sedimentation equilibrium were fitted to a single species model to establish the weight average molar mass (M_{app}) of the protein for each loading concentration.

determined by fitting the observed sedimentation equilibrium distribution curves to a single species model, where the molar mass that gives the best fit for the curve was determined by least squares non linear regression. The solvent density, of 0.99967 g mL^{-1}, was measured using a DMA 5000 densitometer (Anton Paar). The partial specific volume of H-NS_{1-83} C21S was calculated to be 0.7312 ml g^{-1} from the protein sequence, using SEDNTERP (Laue et al., 1992). This change in apparent molar mass indicates an equilibrium between a homodimer (calculated molar mass is 19260 g mol^{-1}) and a homotetramer (calculated molar mass is 38520 g mol^{-1}) of H-NS_{1-83} C21S. Global fitting to a model for a homodimer to homotetramer equilibrium gives a best fit with a dissociation constant of 340 µM (Figure 3-17).

3.12.2 Analysis of the Oligomerisation of H-NS_{1-83} C21S by Analytical Ultracentrifugation in 150 mM $NaClO_4$

Sedimentation velocity data for H-NS_{1-83} C21S in 20 mM MOPS pH 7.0, 150 mM $NaClO_4$ were acquired at 40000 rpm in two sector cells. Direct boundary fitting was performed using SEDPHAT, with the c(s) distribution model for all four protein concentrations: 438 µM, 217 µM, 104 µM and 59 µM (Figure 3-18A). The position of the meniscus and bottom of the cell were determined from their positions observed in 280 nm absorbance scans of the cells. The weight average frictional ratio and distribution of non-interacting species within the sedimentation coefficient range of 0.75 S to 6S was optimised by non linear regression. The c(s) distribution curves are shown in Figure 3-18B.

As the protein concentration is increased from 59 µM to 438 µM H-NS_{1-83} C21S larger species are observed in the c(s) distribution indicating self association of the protein. The concentration dependent increase in the weight average sedimentation coefficient for H-NS_{1-83} C21S is shown in Figure 3-18C. When H-NS_{1-83} C21S was analysed in 20 mM MOPS pH 7.0 without any $NaClO_4$ present the largest species observed in the c(s) distribution plot had a sedimentation coefficient of approximately 2.6 S. After correcting for the increase in solvent density and viscosity that results from the addition of 150 mM $NaClO_4$, using SEDNTERP, this equates to a sedimentation coefficient of 2.39 S. Integration of the c(s) distribution for the range from 0.73 S to 2.39 S shows that at 58 µM H-NS_{1-83} C21S 65 % of the protein corresponds to species within this range of sedimentation coefficients. When the protein concentration is increased to 438 µM only 26 % of the total protein present is found between 0.73 S and 2.39 S in the c(s)

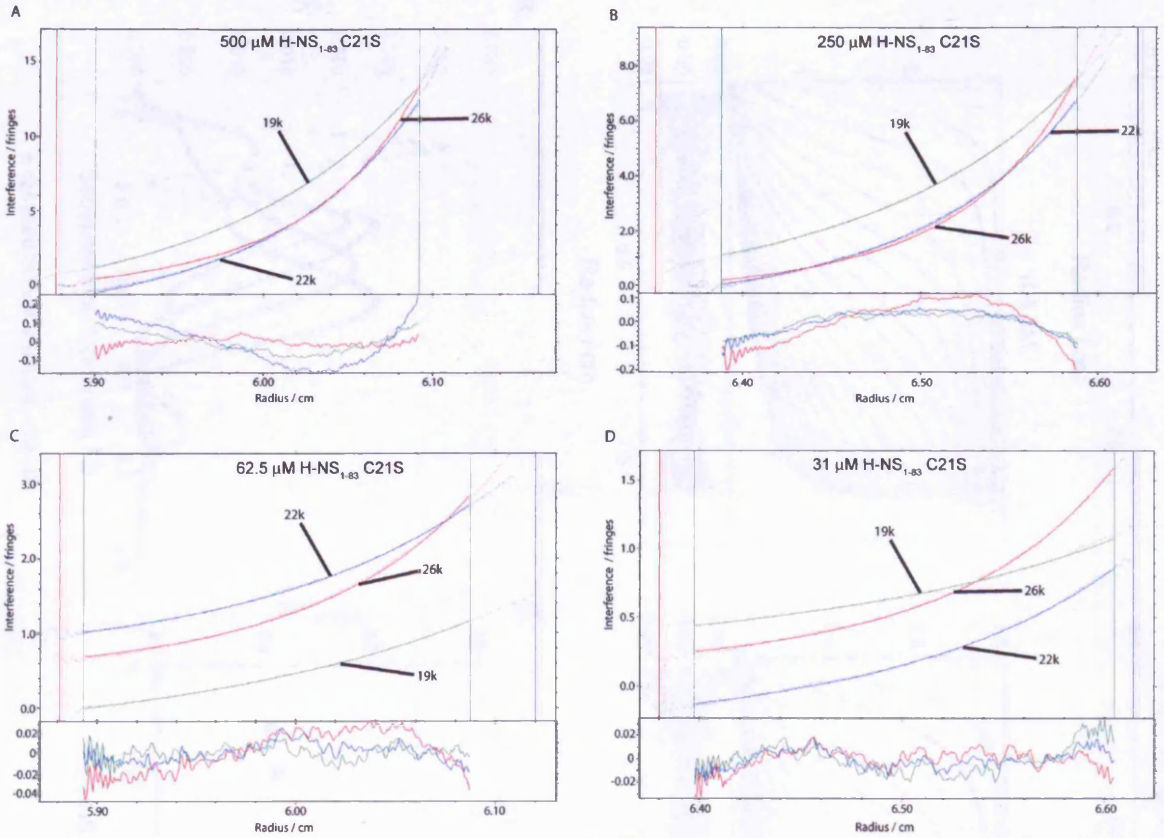


Figure 3-17 Sedimentation equilibrium analytical ultracentrifugation analysis of H-NS_{1-83} C21S in 20 mM Mops pH 7.0. Data was collected for three different rotor speeds; 19000 rpm, 22000 rpm and 26000 rpm for 500 µM, 250 µM, 63 µM and 31 µM starting protein concentrations. All of the data was fitted with a global fit using a monomer-dimer self-association model where the molar mass of the monomer was fixed as 19259.8 g mol^{-1} (the calculated molar mass of an H-NS_{1-83} C21S homodimer). The residuals for the fit are shown below each data set. The best fit was obtained with a dissociation constant of 340 µM with a global reduced Chi squared of 82.9 for the 4068 data points fitted.

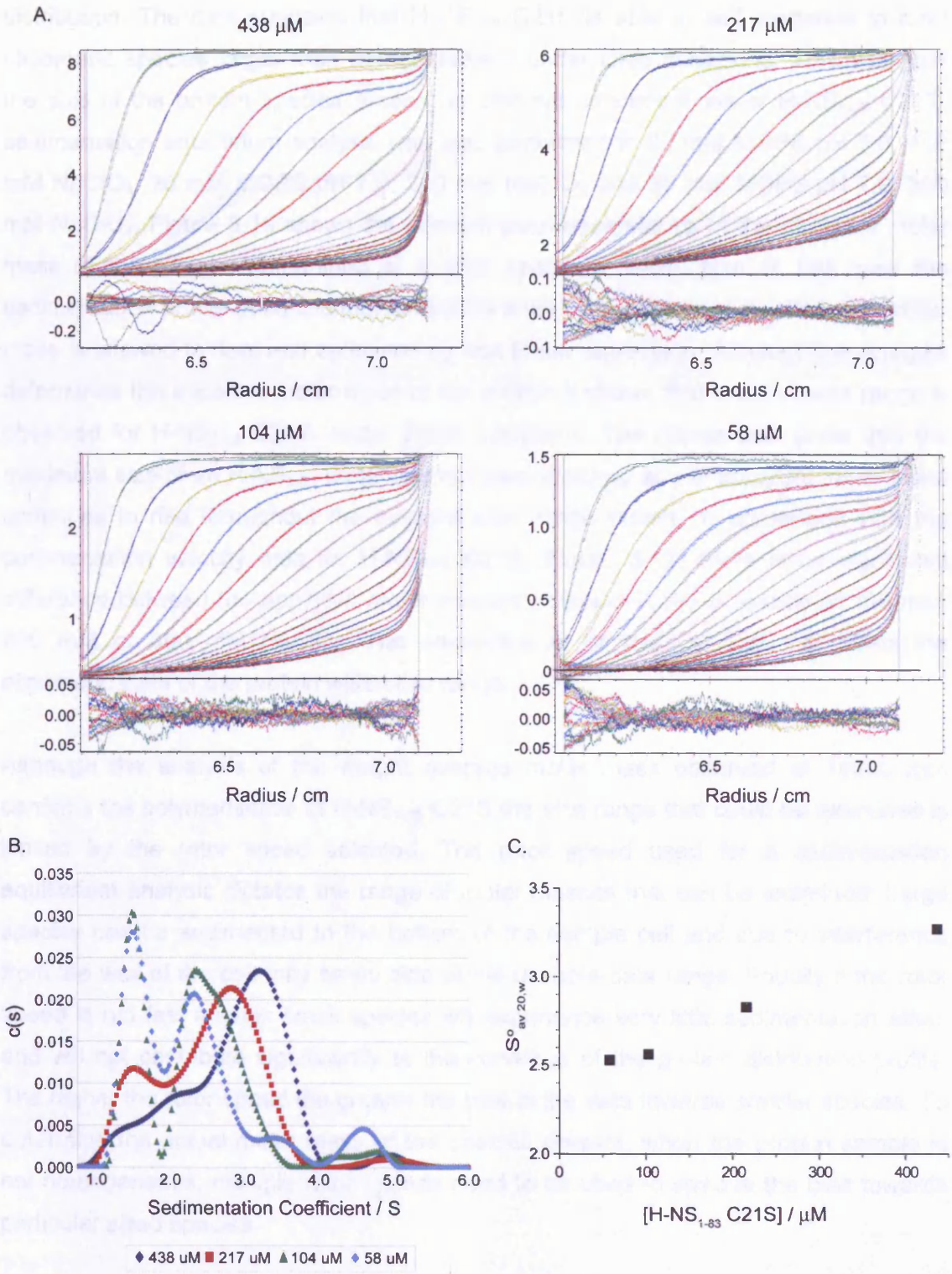


Figure 3-18. Sedimentation velocity analytical ultracentrifugation data for H-NS$_{1-83}$ C21S in 20 mM MOPS pH 7.0, 150 mM $NaClO_4$. A. Sedimentation boundaries fitted using the global c(s) distribution analysis implemented in the SEDPHAT software. The raw data and fitted curves are shown in for each dataset. The residuals for the fitted curves are shown below the raw data in each case. B. The normalised c(s) distribution determined for H-NS$_{1-83}$ C21S at 438 µM, 217 µM, 104 µM and 59 µM loading concentrations. C. The dependence of the weight average sedimentation coefficient on the concentration of H-NS$_{1-83}$ C21S.

distribution. The data suggests that H-NS_{1-83} C21S is able to self associate to form oligomeric species larger than homotetramers under ionic conditions. To investigate the size of the protein species present at different concentrations of H-NS_{1-83} C21S, sedimentation equilibrium analysis was also performed in 20 mM MOPS pH 7.0, 100 mM $NaClO_4$, 20 mM MOPS pH 7.0, 200 mM $NaClO_4$ and 20 mM MOPS pH 7.0, 300 mM $NaClO_4$. Figure 3-19 shows the concentration dependence of the apparent molar mass of the protein, determined at a rotor speed of 14000 rpm. In this case the sedimentation profile at equilibrium is fitted to a single species model, where the molar mass is allowed to float and optimised by non linear regression. Although the analysis determines the apparent molar mass of the protein it shows that a wide mass range is observed for H-NS_{1-83} C21S, under these conditions. The results also show that the maximum size of an H-NS_{1-83} C21S has not been reached as the apparent molar mass continues to rise throughout the concentration range tested. In agreement with the sedimentation velocity data for H-NS_{1-83} C21S (Figure 3-13) there is no significant difference between the apparent molar masses obtained in the presence of 100 mM, 200 mM or 300 mM $NaClO_4$. The concentration of $NaClO_4$ does not affect the oligomeric state of the protein within this range.

Although the analysis of the weight average molar mass observed at 14000 rpm confirms the polymerisation of H-NS_{1-83} C21S the size range that could be examined is limited by the rotor speed selected. The rotor speed used for a sedimentation equilibrium analysis dictates the range of molar masses that can be examined. Large species can be sedimented to the bottom of the sample cell and due to interference from the wall of the cell may be outside of the useable data range. Equally if the rotor speed is not fast enough small species will experience very little sedimentation effect and will not contribute significantly to the curvature of the protein distribution profile. The higher the rotor speed the greater the bias in the data towards smaller species. To determine the actual molar mass of the species present, when the protein sample is not homogeneous, multiple rotor speeds need to be used to remove the bias towards particular sized species.

Sedimentation equilibrium AUC data for five rotor speeds: 8000 rpm, 14000 rpm, 20000 rpm, 26000 rpm and 32000 rpm, for a range of protein sample concentrations from 8 µM to 530 µM have been used to assess the species distribution. This analysis uses the superposition of curves for various species of molar mass corresponding to multiples of H-NS_{1-83} C21S monomers. In the analysis the species are treated as non-interacting species to reduce the number of parameters to be fitted, with the relative proportion of each species optimised to best account for the shape of the raw data

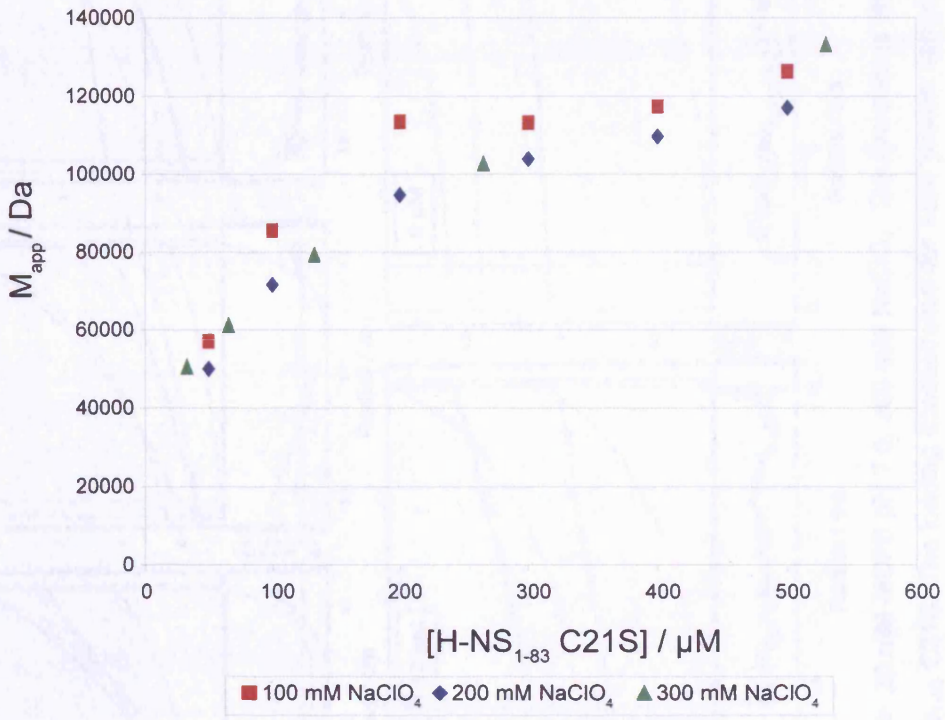


Figure 3-19. Sedimentation equilibrium analysis of various loading concentrations of H-NS_{1-83} C21S prepared in 20 mM MOPS pH 7.0, 100 mM $NaClO_4$ (red), 20 mM MOPS pH 7.0, 200 mM $NaClO_4$ (blue) or 20 mM MOPS pH 7.0, 300 mM $NaClO_4$ (green). Each datapoint is shows the apparent molar mass, determined by fitting the sedimentation equilibrium profile obtained at each loading concentration, to a single species fit where the molar mass was determined by least squares non linear regression. The rotor speed was set at 14000 rpm.

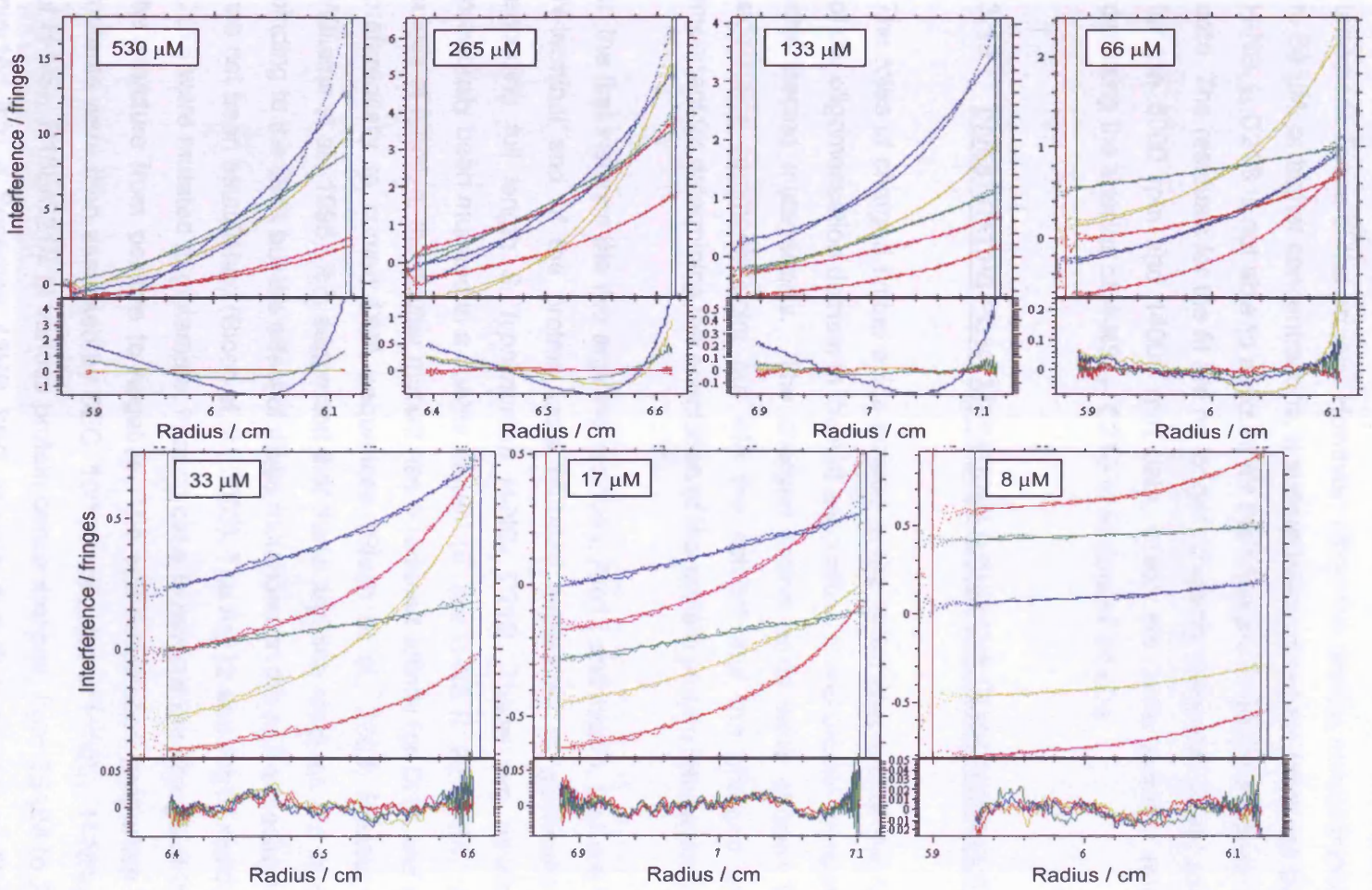


Figure 3-20. Sedimentation equilibrium analysis for H-NS_{1-83} C21S in 20 mM MOPS pH 7.0, 300 mM $NaClO_4$. The raw data is fitted using a superpose curves for non interacting species of monomeric, dimeric and tetrameric H-NS_{1-83} C21S. The loading concentration for each protein sample is indicated. The data for rotor speeds of 8000 rpm (green), 14000 rpm (blue), 20000 (yellow), 26000 rpm (pink) and 32000 rpm (red) are shown. The dots and lines show the raw data and fitted curve respectively.

curve. Figure 3-20 shows the fitting of the sedimentation equilibrium profiles for H-NS_{1-83} C21S in 20 mM MOPS pH 7.0, 300 mM $NaClO_4$ when the molar masses of the species present are limited to monomeric (9629.9 g mol^{-1}), dimeric (19260 g mol^{-1}) and tetrameric (38520 g mol^{-1}) H-NS_{1-83} C21S. The results suggest that between 8 μM and 33 μM H-NS_{1-83} C21S the protein is in equilibrium between monomeric, dimeric and tetrameric forms of the protein. However, when the protein concentration is increased to 66 μM, or higher concentrations, a superposition of curves for these three species of H-NS_{1-83} C21S is not able to account for the changes in curvature observed in the raw data. The residuals for the fit are no longer randomly disributed about zero, particularly for the 8000 rpm and 14000 rpm data, which are more suitable rotor speeds for detecting the species of H-NS_{1-83} C21S in excess of 39 kDa.

3.13 Investigating Point Mutations within the Oligomerisation Domain

The roles of charged amino acids located at the N-terminal end or the C-terminal end of the oligomerisation domain in the self association of the protein were investigated by site-directed mutagenesis. The charged amino acids were chosen because their sidechains commonly point out into the solvent and the charged groups will be important for determining the orientation of the protein:protein interactions.

In the first instance the two arginine residues, Arg12 and Arg15, that are located at the N-terminal end of the protein were mutated separately to glutamate in the DNA encoding full length *S. typhimurium* H-NS_{fl} C21S. These two amino acids have previously been mutated in a double mutant (*E. coli* H-NS R12E/R15A), which showed a loss of function; the double mutant has a reduced affinity for DNA and does not bind preferentially to curved DNA sequences (Bloch et al., 2003; Badaut et al., 2002; Williams et al., 1996). It is suggested that these arginine residues are involved in direct binding to the DNA but the effect of these mutations on the self-association interactions has not been established (Bloch et al., 2003). The Arg12 and Arg15 residues in H-NS_{FL} C21S were mutated to glutamate, in each case to reverse the charge at that position in the structure from positive to negative. The self-association properties of these two mutants were then assessed by SEC. 100 μL samples of H-NS_{FL}, H-NS_{FL} R12E/C21S of H-NS_{FL} R15E/C21S at various protein concentrations, from 25 μM to 200 μM, were injected into a Superose 12HR SEC column and their elution profiles determined (Figure 3-21).

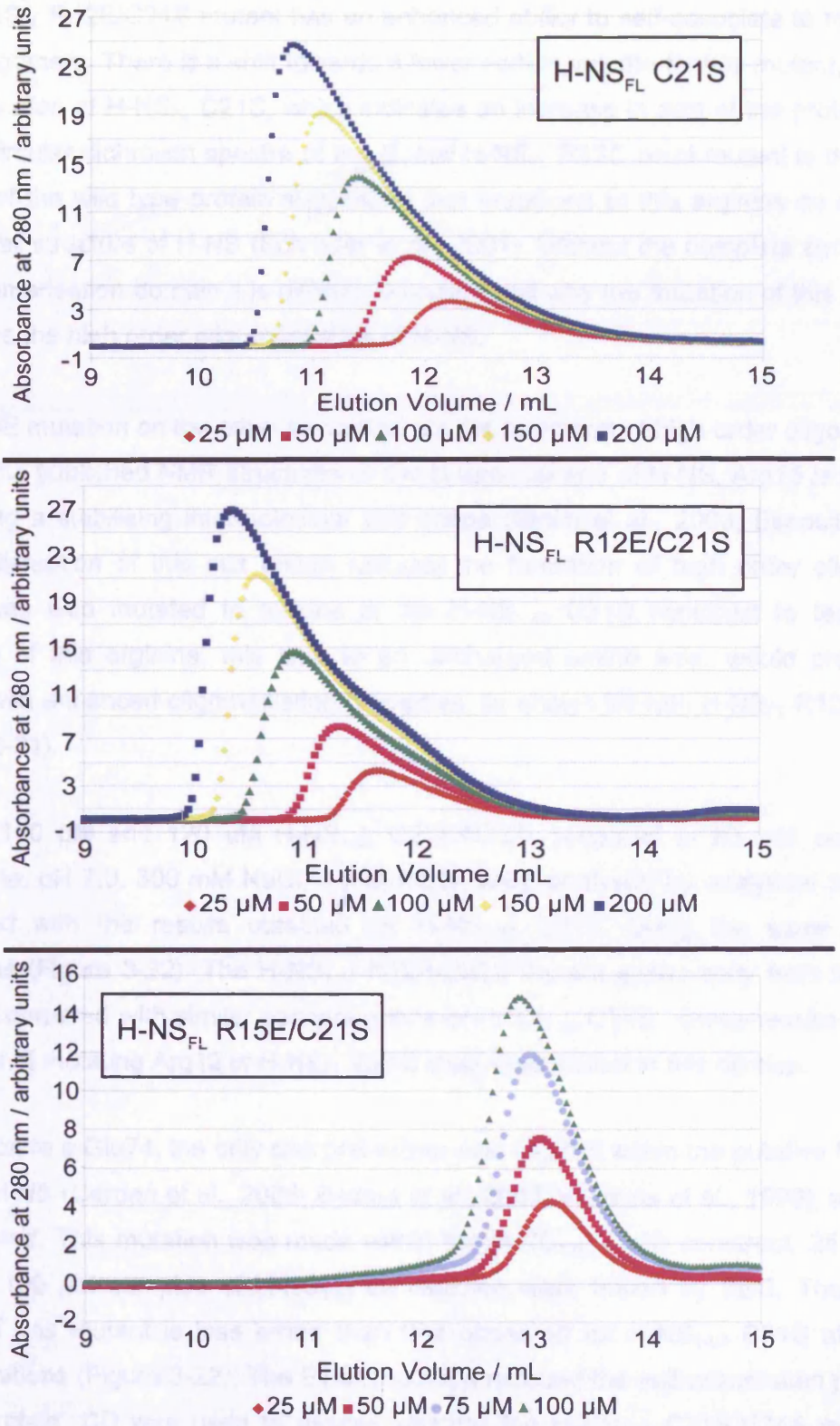


Figure 3-21. Analytical size exclusion chromatography titration of H-NS$_{FL}$ C21S, H-NS$_{FL}$ R12E/C21S and H-NS$_{FL}$ R15E/C21S. The protein samples were dissolved in 20 mM potassium phosphate, pH 7.0, 300 mM NaCl, 1 mM EDTA. 100 µL protein samples at the concentrations shown were injected into a Superose 12HR analytical column (GE Life Sciences) at a flow rate of 0.75 mL min^{-1}.

The H-NS_{FL} R12E/C21S mutant has an enhanced ability to self-associate to form high order oligomers. There is a shift towards a lower elution volume for this mutant, relative to the titration of H-NS_{FL} C21S, which indicates an increase in size of the protein. The far UV circular dichroism spectra of the *E. coli* H-NS_{FL} R12C point mutant is the same as that of the wild type protein suggesting that mutations to this arginine do not alter the overall structure of H-NS (Schroder et al., 2001). Without the complete structure of the oligomerisation domain it is difficult to understand why the mutation of this arginine enhances the high order oligomerisation of H-NS.

The R15E mutation on the other hand disrupts the formation of high order oligomers. In both of the published NMR structures of the N-terminal end of H-NS, Arg15 is involved in forming a stabilising intermolecular salt bridge (Bloch et al., 2003; Esposito et al., 2002); disruption of this salt bridge reduces the formation of high order oligomers. Arg12 was also mutated to alanine in the H-NS_{1-83} C21S construct to test if the mutation of this arginine, this time to an uncharged amino acid, would produce a mutant with enhanced oligomeisation properties, as shown for with H-NS_{FL} R12E/C21S (Figure 3-21).

50 µM, 100 µM and 120 µM H-NS_{1-83} C21S/R12A, prepared in 20 mM potassium phosphate, pH 7.0, 300 mM NaCl, 1 mM EDTA were analysed by analytical SEC and compared with the results obtained for H-NS_{1-83} C21S, using the same solvent conditions (Figure 3-22). The H-NS_{1-83} R12A/C21S mutant elutes early from the SEC column compared with similar concentrations of H-NS_{1-83} C21S. These results confirm the effect of mutating Arg12 in H-NS_{FL} C21S discussed earlier in this section.

A mutation to a Glu74, the only charged amino acid located within the putative fourth α-helix of H-NS (Cerdan et al., 2003; Badaut et al., 2002; Williams et al., 1996), was also investigated. This mutation was made within the H-NS_{1-83} C21S construct. 25 µM, 50 µM and 100 µM samples of H-NS_{1-83} C21S/E74A were tested by SEC. The elution profile of this mutant is less broad than that observed for H-NS_{1-83} C21S at similar concentrations (Figure 3-22). The E74A mutation reduced the self-association potential of the protein. CD was used to assess whether the H-NS_{1-83} C21S/E74A mutant is partially unfolded by comparing the far-UV CD sprecta of these two proteins. Both protein samples were prepared in 20 mM potassium phosphate, pH 7.0, 300 mM NaCl, 1 mM EDTA. The far-UV CD spectra recorded for these two proteins are overlaid in Figure 3-23B. The far-UV spectrum of H-NS_{1-83} C21S/E74A shows slightly less structure (less negative mean residue ellipticity at 208 nm and 222 nm) when compared with the CD spectrum for H-NS_{1-83} C21S although the results are within the

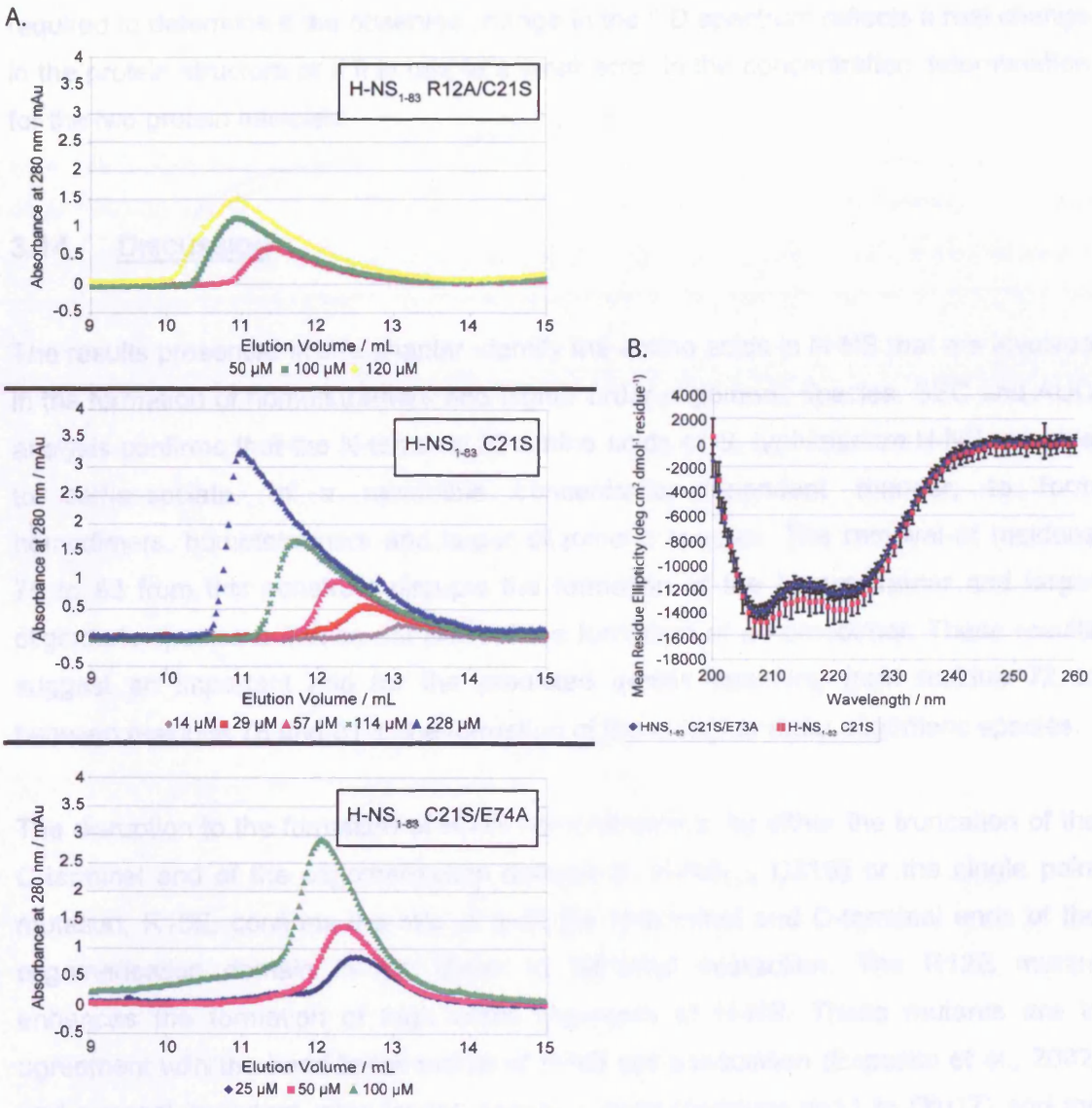


Figure 3-22 A. Analytical size exclusion chromatography titration of H-NS_{1-83} R12A/C21S, H-NS_{1-83} C21S and H-NS_{1-83} C21S/E74A. The protein samples were dissolved in 20 mM potassium phosphate, pH 7.0, 300 mM NaCl, 1 mM EDTA. 100 µL protein samples at the loading concentrations shown were injected into a Superose 12HR analytical column (GE Life Sciences) at a flow rate of 0.75 mL min^{-1}. B. The overlaid circular dichroism spectra for H-NS_{1-83} C21S/E74A and H-NS_{1-83} C21S. Spectra between 260 nm and 195 nm were recorded. Error bars show the variation of the circular dichroism signal observed over a five second averaging time for each wavelength.

error determined from the variation in the CD signal. Further experimentation is required to determine if the observed change in the CD spectrum reflects a real change in the protein structure or if it is due to a small error in the concentration determination for the two protein samples.

3.14 DISCUSSION

The results presented in this chapter identify the amino acids in H-NS that are involved in the formation of homotetramers and higher order oligomeric species. SEC and AUC analysis confirms that the N-terminal 83 amino acids of *S. typhimurium* H-NS are able to self-associate, in a reversible concentration-dependent manner, to form homodimers, homotetramers and larger oligomeric species. The removal of residues 75 to 83 from this construct disrupts the formation of the homotetramer and larger oligomeric species but does not prevent the formation of a homodimer. These results suggest an important role for the predicted α-helix spanning from residue 72 to between residues 78 and 81 in the formation of these higher order oligomeric species.

The disruption to the formation of H-NS homotetramers, by either the truncation of the C-terminal end of the oligomerisation domain (in H-NS_{1-74} C21S) or the single point mutation, R15E, confirms the role of both the N-terminal and C-terminal ends of the oligomerisation domain in the dimer to tetramer interaction. The R12E mutant enhances the formation of high order oligomers of H-NS. These mutants are in agreement with the head-to-tail model of H-NS self-association (Esposito et al., 2002) and suggest important roles for the second α-helix (residues Ile11 to Glu17) and the predicted fourth α-helix in the polymerisation process.

The DNA binding domain is not required for the formation of tetramers and high order oligomers, in agreement with previous observations for H-NS_{1-90} C21S and analysis of the full length H-NS protein by NMR (Shindo et al., 1995; Smyth et al., 2000). However, other studies have led to the conclusion that the DNA binding domain of H-NS is able to form homodimers and mediates the formation of tetramers and higher order oligomeric species (Spurio et al., 1997; Stella et al., 2005). The principal experiments in these two publications rely on the use of fusion proteins, where H-NS is fused to a sequence specific DNA binding domain from two phage repressor proteins called P22cI and 434cI. The repression of the reporter gene required the binding of multiple DNA binding domains to repress gene expression as the binding of either P22cI or 434cI alone was not sufficient to inhibit expression of the β-galactosidase reporter gene. The

level of reporter gene repression by the fusion protein with H-NS(Δ1-89)ΔP116 was reduced compared with the level observed with the wild type H-NS fusion protein. The H-NS(Δ1-89)ΔP116 protein contains the C-terminal nucleic acid binding domain of H-NS where proline 116 has been deleted. The reduced repression of the reporter gene was interpreted as a disruption to the high order oligomerisation of H-NS (Spurio et al., 1997; Stella et al., 2005) However, these results can also be attributed to the cooperative binding of both H-NS and the sequence specific DNA binding protein to the DNA. Appropriate controls were not prepared to demonstrate that a construct containing the nucleic acid binding domain of H-NS fused to the sequence specific DNA binding domain (either P22cI or 434cI) do not act cooperatively to repress the reporter gene expression without the need for dimerisation. If the H-NS nucleic acid binding domain binds to the DNA when fused to the P22cI DNA binding domain and enhances the overall DNA binding interaction, the reduced repression of the reporter gene when the H-NS ΔP116 mutant is expressed can be explained by a reduced affinity of the H-NS nucleic acid binding domain for DNA. Proline 116 is located within the highly conserved DNA binding loop which has been shown to be partially responsible for the DNA binding of H-NS (Shindo et al., 1999). Mutation of proline 116 to alanine or serine reduces the affinity of H-NS for DNA (Badaut et al., 2002).

The interaction of H-NS_{1-83} C21S homodimers to form larger oligomeric species was disrupted by reducing the concentration of $NaClO_4$ in the buffer below 100 mM. Sedimentation velocity and sedimentation equilibrium AUC experiments established that an equilibrium between homodimer and homotetramer exists, when no $NaClO_4$ was present in the solvent, confirming the existence of a homotetramer. Increasing the $NaClO_4$ concentration of the buffer, from zero to 100 mM $NaClO_4$, enhances the self-association interactions allowing larger oligomeric species to form. SEC analysis of the apparent molar mass of the H-NS_{1-83} C21S, in the presence and absence of 300 mM NaCl, demonstrates that the oligomerisation interactions are enhanced by the addition of NaCl. The N-terminal domain of H-NS is predominantly negatively charged; the theoretical pI of H-NS_{1-83} C21S is 4.72 (Bjellqvist et al., 1993) and H-NS_{1-83} C21S binds to an anion exchange chromatography column at pH 7.0. These results suggest that the addition of Na^+ ions shields repulsion forces between H-NS_{1-83} C21S homodimers, increasing the potential for larger oligomers to form.

The distribution of oligomeric species was only affected by changes in the $NaClO_4$ concentration of the solvent at concentrations below 100 mM. When varying concentrations from 100 mM to 300 mM $NaClO_4$ were examined by AUC the distribution of species was independent of the $NaClO_4$ concentration. In contrast, large

zone size exclusion chromatography studies showed that the elution volume of H-NS increases when 200 mM NaCl was replaced with 200 mM NH_4Cl, suggesting that the formation of tetramers was disrupted by the change from Na^+ to NH_4^+ ions (Ceschini et al., 2000). It is unclear whether the change observed in the large zone chromatography experiments can be attributed to changes in the strength of the interaction between the protein and the column matrix when the solvent is altered. The Superose 12HR calibration curves produced in the presence and absence of 300 mM NaCl shows a significant shift in the elution volume of the globular proteins used to calibrate the column (Figures 2-15 and 2-16), showing that changes to the salt concentration of the buffer can affect the elution volume observed.

Changes in *E. coli* cytoplasmic ion concentrations observed in response to changes in the extracellular environment are much greater than the range of NaCl or $NaClO_4$ concentrations examined in this research. Potassium and glutamate ions in the cytoplasm are increased to prevent dehydration of the cell and maintain cell turgidity (Leirmo et al., 1987). The potassium ion concentration increases from 0.2 to 0.9 molal (moles of solute per kilogram of solvent), in response to a change in osmolality of the growth media from 0.1 molal to 1.1 molal. For the same change in the external environment the intracellular glutamate concentration also rose from 0.03 to 0.25 molal (Richey et al., 1987). Further research is required to examine the effect of much higher salt concentrations than those investigated in this chapter on the high order oligomerisation interactions of H-NS.

The intracelluar concentration of H-NS in the *E. coli* cytoplasm is estimated at approximately 20 µM (Bouffartigues et al., 2007). The sedimentation equilibrium AUC analysis of H-NS in 20 mM MOPS pH 7.0, 300 mM $NaClO_4$ shows that H-$NS_{1\text{-}83}$ C21S is in equilibrium between a monomer, dimer and a tetramer at this concentration. Whilst the importance of forming homodimers, which have a higher affinity for DNA than the single nucleic acid binding domain (Shindo et al., 1999) is known, the role of H-NS tetramers and larger oligomeric species in the cell is not fully understood. DNase I footprinting studes for gene promoters that are known to be repressed by H-NS has identified multiple H-NS binding sites close to the transcription start site (Bouffartigues et al., 2007; Falconi et al., 1993; Falconi et al., 1998; Lucht et al., 1994; Prosseda et al., 2004). Quantitative analysis of the protection of the DNA against DNase I digestion by increasing H-NS concentrations, shows different affinities of H-NS for binding sites within the DNA (Bouffartigues et al., 2007). Two models for H-NS mediated gene repression have emerged from these studies.

In one model H-NS binds to a high affinity recognition site and recruits more H-NS molecules which polymerise along the DNA, forming large oligomeric species of H-NS that block transcription. In the other model, H-NS bridges between two H-NS binding regions, bending the DNA back upon iteslf in a hairpin like structure. These models are discussed in greater detail in section 1.10. The formation of DNA bridges is required for H-NS mediated compaction of the DNA, as observed by atomic force microscopy (Dame et al., 2000). Whilst the first model requires the formation of tetramers and larger oligomeric species the second model could be explained by DNA bridging from multiple H-NS homodimers, although it does not preclude the binding of larger oligomeric species. Experiments using optical tweezers to manipulate H-NS bridged DNA molecules (Figure 3-23) bring into question the role of tetramers and larger species of H-NS in the function of the protein (Dame et al., 2006; Dorman, 2007a). A force was applied to unzip the bridged DNA molecules pulling the DNA duplexes apart in steps whose size depends on the spacing between H-NS bridges. The smallest step observed corresponds to a single turn of a DNA helix (3.5 nm) leading to the conclusion that homodimers of H-NS bridge between the two DNA duplexes. Tetramers of H-NS and larger oligomers were not identified in the analysis suggesting that the cooperative binding of H-NS homodimers to DNA is brought about by the proximity of the two DNA duplex molecules to each other. When the DNA strands are close enough for an H-NS homodimer to bridge between the two DNA duplexes other H-NS homodimers will be recruited to make additional bridges between the DNA molecules. The results did not show the presence of tetramers or larger H-NS complexes.

However, a model where H-NS homodimers cooperatively bridge between two regions of DNA fails to account for the effect of point mutations within the oligomerisation domain of H-NS that disrupt the formation of tetramers and higher order oligomers but do not prevent homodimer formation. Chemical crosslinking studies, for both H-NS L26P and H-NS E53G/T55P demonstrate the ability of these mutants to form homodimers. H-NS L26P has the same sedimentation coefficient as the H-NS wild type homodimer in a sedimentation velocity AUC experiment (Badaut et al., 2002). The H-NS L26P and H-NS E53G/T55P mutants bind to high affinity H-NS binding sites in proU promoter DNA, but are unable to bind to the lower affinity H-NS binding sites that are observed when the H-NS concentration is increased above 1 μM. These lower affinity sites require the formation of tetramers and possibly larger species for H-NS to bind and may play an important role the dynamic control of gene expression in *E. coli*.

Figure 3-23. Manipulation of H-NS bridged DNA molecules using optical tweezers. Biotin labelled DNA is bound to streptavidin coated beads (1-4) which can be manouvered into the desired positions for teh experiment. One bead (1) is anchored whilst a force is applied to bead 2 to pull the two DNA molecules apart. This Figure is adapted from (Dame et al., 2006).

4 CRYSTALLISATION OF THE OLIGOMERISATION DOMAIN OF H-NS

4.1 INTRODUCTION

In this chapter the systematic approach taken to develop a crystallisation method for the H-NS oligomerisation domain and the techniques used to enhance the resolution of the protein crystals produced are described. The formation of a range of different oligomeric species in a reversible concentration dependent manner, as demonstrated in the previous chapter, provides a significant challenge to a structural biologist. In order to solve a protein structure by NMR or X-ray crystallography it is desirable to have a homogeneous protein population.

4.2 CRYSTALLISATION

In principle the X-ray scattering from a single molecule contains all the information about the protein structure, however the intensity of the scattering signal cannot be detected above the background noise. If the orientation of the molecule could be fixed and the X-ray scattering measured enough times it would be possible to generate sufficient signal intensity above the background, to detect the complete X-ray scattering pattern and thus solve the protein structure. However, protein molecules are often too small to manipulate in single molecule techniques and repeated exposure to an X-ray beam damages the protein molecule. To overcome this problem it is necessary to produce protein crystals where large numbers of protein molecules are arranged in a regular crystalline lattice. When an X-ray beam is scattered by the molecules in the crystal constructive and destructive interference occurs between the scattered X-rays and a diffraction pattern can be detected. Each spot on the diffraction pattern contains information about the relative position of all of the atoms in the molecule. The growth of protein crystals suitable for structure determination is largely a matter of trial and error and provides a significant bottleneck in the process of structure determination.

4.2.1 The Formation of Protein Crystals

A crystal is a highly ordered array of molecules where the relative position of each molecule to another is consistent throughout the crystal lattice. Each repeating unit, called the unit cell, must be translated or rotated in a consistent way throughout the crystal in order to generate sufficient signal to noise in the diffraction pattern detected to be able to solve the protein structure. The difficulty arises from the challenge of producing a regular crystal lattice from the normally irregularly shaped protein molecules.

In order to produce a protein crystal a solution of the molecule of interest needs to be prepared and a precipitant is added to decrease the solubility of the protein, so that it starts to fall out of solution. Adding too much precipitant will cause the protein to precipitate in a disordered manner. However, under certain conditions the protein molecules will come together in an ordered manner to form the beginnings of a crystal, called nucleation. Once the nucleus of a crystal is formed protein molecules will continue to add to the crystal as long as the protein in solution is at a supersaturated concentration. There is a significant energy barrier in the formation of crystals but once overcome it is energetically favourable for the protein molecules in solution to bind to the already nucleated crystal, as opposed to forming other nuclei themselves. The protein concentration, or alternatively the precipitant concentration required to achieve nucleation is greater than the concentration needed to build up on an already formed crystal nuclei. Under conditions where a crystal will grow but not nucleate the protein solution is said to be in a metastable state. A phase diagram showing the effect of changing the protein or precipitant concentration on the solubility of the protein is shown in Figure 4-1.

Several different methods have been developed to create conditions where the precipitant concentration gradually increases over time, in order to identify the point required for nucleation of a crystal (Figure 4-2). The most common method is vapour diffusion using either a hanging drop or sitting drop setup. In a vapour diffusion arrangement a small volume of solution containing the protein of interest and a solvent containing precipitant (and buffer components) is either hung from a glass cover slip (hanging drop) or placed on a raised platform (sitting drop) above a reservoir containing the precipitant solution in a sealed volume. As the precipitant concentration in the drop is lower than that found in the reservoir, the water in the protein drop will diffuse out of the drop resulting in a gradual increase in the protein and precipitant

Figure 4-1. Crystallisation phase diagram. The arrows show the transition of successful crystallisation conditions using the vapour diffusion, dialysis and microbatch methods where nucleation of a crystal leads to crystal growth by transition to the metastable phase. Diagram has been adapted from (Bergfors, 1999).

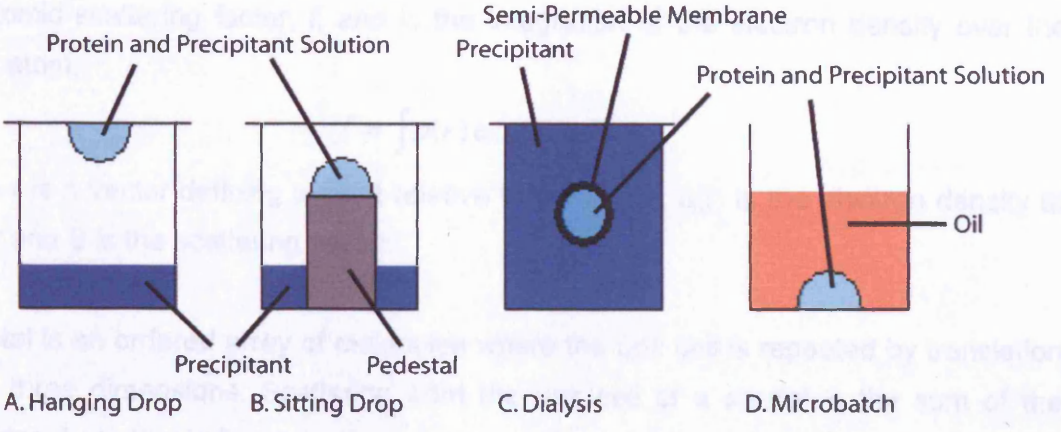


Figure 4-2. Crystallisation trial setup. Diagrammatic representation of hanging drop, sitting drop, dialysis and microbatch methods for the preparation of protein crystals.

concentrations. Two other crystallisation trial setups, microbatch and microdialysis, are commonly used to produce protein crystals.

Microbatch works by adding the precipitant to the protein solution under oil in an attempt to reach nucleation conditions directly without the need for slow diffusion. As the protein crystal grows the concentration of protein in solution within the drop will decrease so the solution will change from a condition suitable for nucleation to a metastable condition. The volume of the drop may decrease over time if an oil is selected that allows solvent to diffuse through it, to allow changes in the protein and precipitant concentration to occur.

In the dialysis method the protein solution is separated from the precipitant solution by a semi-permeable membrane allowing the slow diffusion of precipitants into the protein solution. The precipitant concentration will increase over time to induce nucleation of a protein crystal if the appropriate conditions are met. Then the reduced protein concentration in solution, as the protein crystal grows, will move the condition into the metastable zone if large crystals are to be obtained.

4.3 X-RAY SCATTERING

As X-rays pass through matter their fluctuating electric field interacts with the electrons causing them to oscillate. When the electrons return to their original energy state they emit a photon of the same wavelength as the incident X-ray but in a random direction. This process is called elastic scattering. The total scattering by an atom is known as the atomic scattering factor, *f*, and is the integration of the electron density over the whole atom.

$$f = \int \rho(r) \exp(2\pi i r S) dr$$

where r is a vector defining a point relative to the origin, ρ(r) is the electron density at point r and S is the scattering vector.

A crystal is an ordered array of molecules where the unit cell is repeated by translation in the three dimensions. Scattering from the unit cell of a crystal is the sum of the scattering from the individual atoms in the unit cell. The structure factor, *F*(S) for a wave scattered from the entire unit cell can be calculated from the integration of the electron density of the entire unit cell.

$$F(S) = \int_{cell} \rho(r) \exp(2\pi i r S) dr$$

Interference will occur between X-rays scattered from the individual atoms within the unit cell to produce a unique diffraction pattern for the unit cell. A crystal replicates the X-ray scattering from each unit cell in order to amplify the signal, allowing the diffraction pattern to be detected.

4.4 DIFFRACTION

The diffraction pattern is produced by interference of the scattered X-rays. When two waves with the same wavelength interfere they combine to form a wave with the same wavelength but with amplitude and phase that is the product of the two individual waves (Figure 4-3). If the scattered X-rays are in phase then constructive interference will occur leading to a spot on the diffraction pattern. If the X-rays are out of phase then destructive interference will occur resulting in a blank region in the diffraction pattern. Each reflection or spot in the diffraction pattern will contain some information about all of the atoms in the unit cell. The spots, or reflections, in the diffraction pattern occur because of constructive interference between X-rays scattered from different unit cells within the crystal that are located in parallel planes.

William Henry Bragg together with William Laurence Bragg showed that the X-ray diffraction from a crystal can be thought of as reflections from a series of parallel planes running through the crystal. In this analogy when an X-ray beam enters a crystal some of the X-rays will be reflected by atoms in the first layer of the crystal lattice whilst other X-rays will penetrate this layer and be reflected by the next layer of atoms. The reflected rays will only be in phase when the angle of incidence of the X-ray beam satisfies Bragg's law

$$n\lambda = 2d \sin\theta$$

where n is an integer number, λ is the wavelength of the incident X-ray beam, d is the distance between the atoms in the lattice plane and θ is the angle of incidence for the X-ray beam.

When the distance travelled by two reflected X-rays differs by an integer number of wavelengths then the emitted X-rays are said to be in phase. The distance travelled by two X-rays reflected from different parallel layers in the crystal lattice is shown in (Figure 4-4).

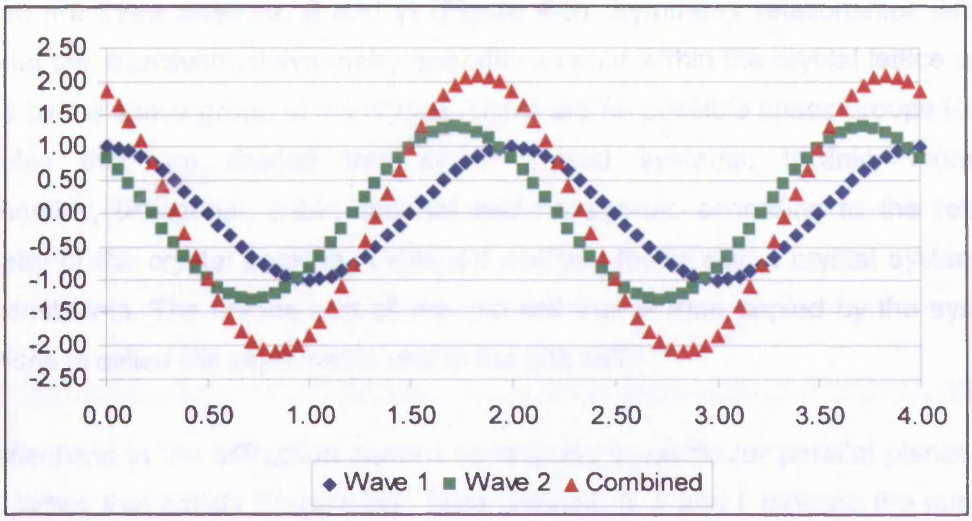


Figure 4-3. Interference of two waves of equal wavelength but different amplitude and phase. The individual waves are shown in blue and green respectively. The combined interference wave is shown in red. The combined wave has the same wavelength as the individual waves but its amplitude and phase is the product of the two individual waves.

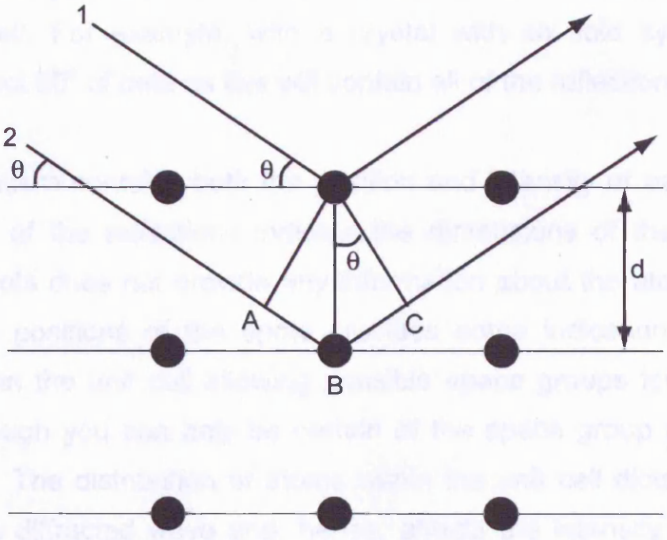


Figure 4-4. Reflection of X-rays from the first and second row of atoms in a crystal lattice. The extra distance travelled by the reflection of incident ray 2 from the second row of atoms is the sum of the distances AB and BC. AB+BC = 2dsinθ, where θ is the angle of incidence and d is the distance between the Bragg planes.

The unit cell is characterised by the length of each side (a, b and c) and the angles between the three sides (α, β and γ) (Figure 4-5). Symmetry relationships with both rotational and translational symmetry operations occur within the crystal lattice and are defined by the space group of the crystal. There are 65 possible space groups for chiral molecules that are divided into seven crystal systems: triclinic, monoclinic, orthorhombic, tetragonal, cubic, trigonal and hexagonal, according to the rotational symmetry of the crystal packing. Table 4-1 outlines these seven crystal systems and their constraints. The unique part of the unit cell that is then copied by the symmetry operations is called the asymmetric unit of the unit cell.

The reflections in the diffraction pattern correspond to particular parallel planes of the crystal lattice that satisfy Bragg's law. Miller indices, *h*, *k* and *l*, indicate the number of times the particular set of parallel Bragg planes intersect the A, B and C edges of the unit cell, therefore describing the orientation of the plane relative to an origin position. To determine the structure of a protein the crystal is rotated and the diffraction pattern collected in segments of 0.1° to 1° rotations of the crystal. As the crystal is rotated the planes in the crystal that satisfy Bragg's law will change so different reflections become visible, whilst other reflections disappear. It is not normally required to collect data for all 360° of possible crystal orientations because of the crystallographic symmetry found within the unit cell. For example, with a crystal with six fold symmetry it is only necessary to collect 60° of data as this will contain all of the reflections from the crystal.

The diffraction pattern contains both the position and intensity of each reflection. The relative positions of the reflections indicate the dimensions of the unit cell but the spacing of the spots does not provide any information about the atoms within the unit cell. The relative positions of the spots provides some indication of the symmetry relationships within the unit cell allowing possible space groups for the crystal to be determined, although you can only be certain of the space group once the structure has been solved. The distribution of atoms within the unit cell dictates the amplitude and phase of the diffracted wave and, hence, affects the intensity of the spot in the diffraction pattern. However as the detector is only able to measure the energy of the wave, to give the intensity of the spot, it is not possible to determine the phase of the wave and this information is lost. The intensity of the spots also contains significant errors due to fluctuations in the incident X-ray beam intensity, changes in the volume of the crystal in the X-ray beam as the crystal is rotated, decay of the crystal due to radiation damage and the affects of the cryoprotectant or the fibre loop that holds the crystal on and the intensity of the X-rays reaching the protein molecules.

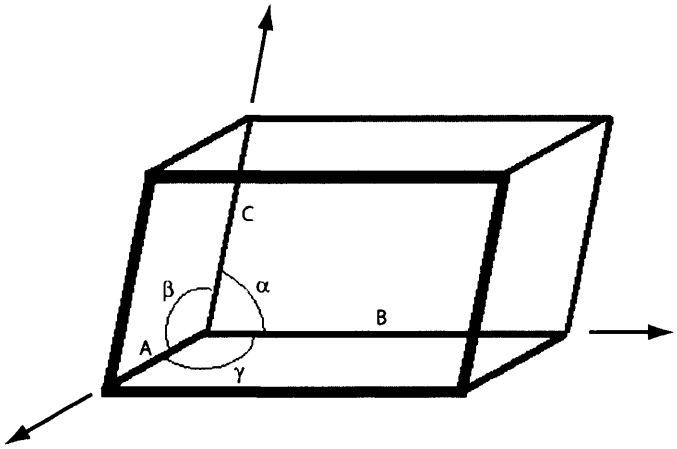


Figure 4-5. A diagram of the unit cell showing the dimensions A, B and C and the angles α, β and γ.

Table 4-1. The seven crystal systems showing their rotational symmetry relationships and the unit cell dimension and angle constraints for each system.

Crystal System	Minimum Symmetry	Constraints
Triclinic	None	None
Monoclinic	One 2-fold (along B)	$\alpha = \gamma = 90^\circ$
Orthorhombic	Three 2-fold	$\alpha = \beta = \gamma = 90^\circ$
Trigonal	3-fold (along C)	A = B, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$
Tetragonal	4-fold (along C)	A = B, $\alpha = \beta \gamma = 90^\circ$
Hexagonal	6-fold (along C)	A = B, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$
Cubic	Four 3-fold axis	A = B =C, $\alpha = \beta = \gamma = 90^\circ$

ELECTRON DENSITY CALCULATION

The electron density at the fractional coordinates of the unit cell, ρ(xyz), can be calculated from the structure factor, *F*(hkl), for each of the reflections using a Fourier transform.

$$\rho(xyz) = \frac{1}{V}\sum_{h}^{\infty}\sum_{k}^{\infty}\sum_{l}^{\infty} F(hkl)\exp[-2\pi i(hx+ky+lz)]$$

where V is the volume of the unit cell. If all of the structure factors are known then the electron density for the whole unit cell can be calculated. However the structure factor is a complex number consisting of a structure factor amplitude, |*F*(hkl)|, and a phase angle, α(hkl).

$$F(hkl) = |F(hkl)|\exp i\alpha(hkl)$$

Therefore in order to calculate the electron density from the unit cell both the amplitude and phase angle for the structure factor corresponding to each reflection in the diffraction pattern needs to the determined.

$$\rho(xyz) = \frac{1}{V}\sum_{h}^{\infty}\sum_{k}^{\infty}\sum_{l}^{\infty}|F(hkl)|\exp[-2\pi i(hx+ky+lz)+i\alpha(hkl)]$$

The intensity, I(hkl), of a reflection is the square of the structure factor amplitude for that reflection and can therefore be calculated. However, the phase angle of the structure factor cannot be experimentally measured, so the electron density cannot be determined directly from the diffraction pattern.

4.5 SOLVING THE PHASE PROBLEM

To determine the position of the electrons within the unit cell and solve the protein structure it is necessary to determine the relative phase angles of the structure factors for all of the reflections. This can be done mathematically if the data is of high resolution, better than 1.2 Å resolution, but more usually the phases need to be predicted by molecular replacement (Lattman, 1985) or determined by further experiments.

4.5.1 Experimental Phasing

To determine the phases experimentally it is necessary to use techniques that will change the intensity of the reflections generated from crystals of the same protein in

the same space group symmetry arrangement. These experiments use a reference wave of known phase to generate interference that results in changes in the intensity of the reflections observed. The two main types of interference phasing experiments used are isomorphous replacement (Watenpaugh, 1985) and anomalous dispersion (Hendrickson *et al.*, 1985).

Isomorphous replacement experiments use heavy atoms such as mercury or uranium, that strongly scatter X-rays because of their high electron density, to create the reference wave. The difference between the diffraction patterns in the presence and absence of the heavy atoms in the crystal can be used to solve the phase problem. The position of the heavy atoms in the unit cell needs to be determined before the electron density for the whole unit cell can be calculated (Watenpaugh, 1985).

For anomalous dispersion the reference wave comes from atoms within the structure that resonate with the wavelength of the radiation used. Commonly selenium atoms are introduced in the form of selenomethionine, or the weak anomalous scattering from the sulphur atoms of cysteine and methionine residues, are used for these experiments (Hendrickson *et al.*, 1985). It is often necessary to collect datasets from multiple crystals using both isomorphous replacement and anomalous scattering to solve the protein structure experimentally.

4.5.2 Molecular Replacement

Molecular replacement uses a known structure or model to orient and position the molecule within the unit cell, so that it is aligned in the same orientation as with the actual molecule in the crystal (Lattman, 1985). The phase angles from the molecular replacement model can be calculated for this arrangement of atoms in the unit cell, and these phases angles used together with the observed structure factor amplitudes, determined from the reflections in the diffraction pattern to calculate the electron density map. X-ray crystallography protein structures are generally the best models to use for molecular replacement as they have well defined positions for the atoms and are usually of a higher resolution than protein structures determined by NMR spectroscopy. Molecular replacement requires a close structural similarity between the known protein and the target protein, usually requiring at least 30% sequence identity between the two proteins. The model may also need to be edited to remove flexible regions or parts of the model that do not match with the target protein due to insertions or deletions.

If the electron density map can be determined for the diffraction pattern observed, then a model of the protein can be built within this electron density map, using knowledge of the protein sequence, main chain and sidechain torsion angle restraints and bond length restraints, to find the best match of the protein structure with the electron density. The model can then be refined to improve the agreement between the calculated structure factors predicted from the model and those observed experimentally.

4.6 Selecting the Appropriate Protein Constructs for X-ray Crystallography

We wanted to solve the structure of the complete oligomerisation domain of H-NS to get an understanding of how dimers interact. The choice of protein construct for X-ray crystallography is critical to the likelihood of producing a crystal that is uniformly packed throughout the crystal and therefore, diffracts to a sufficiently high enough resolution suitable for structure determination. It is important that the protein forms a rigid structure; removal of flexible loops and isolation of individual domains is often required for crystals to grow (Cheng et al., 2003; Musacchio et al., 1994). The modification of sidechains by side-directed mutagenesis, replacing the long flexible hydrophilic amino acids, such as arginine, with alanine residues to reduce local conformational entropy that reduces the uniformity of the crystal packing, has also been used to improve the resolution of crystals (Derewenda, 2004; Watenpaugh, 1985).

One other factor that is critical for the crystallization of a protein is the homogeneity of the protein sample within the crystallization condition. Impurities of other proteins or the presence of multiple oligomeric states of the same protein can "poison" the crystal preventing growth of the crystal or causing local imperfections in the crystal lattice that limit the diffraction resolution. Dynamic light scattering studies have shown that protein crystals are more likely to be produced when the protein sample is homogeneous, (D'Arcy, 1994; Ferre-D'Amare and Burley, 1994), although it is also possible that the crystallisation process will purify out one form of the protein from a mixed population.

The ability of H-NS to form a range of different oligomeric species, as demonstrated in section 3.9, provides a significant challenge to the structural biologist. Full length H-NS protein is unlikely to crystallise due to the flexibility of the linker between the oligomerisation and nucleic acid binding domains (Shindo et al., 1995; Smyth et al., 2000). It may be possible to restrict the relative motion of the domains by binding the

protein to a piece of DNA long enough to bind two DNA binding domains. However the lack of DNA sequence specificity potentially introduces a source of heterogeneity in the protein:DNA complex as alternative binding sites on the DNA molecules may be used. Also, there is the possibility of forming a range of different H-NS:DNA complexes as different H-NS oligomeric species bridge between the DNA duplexes.

An alternative approach is to produce protein crystals for the isolated oligomerisation domain where the flexible linker and nucleic acid binding domain have been removed. The NMR, size exclusion chromatography and analytical ultracentrifugation results presented in the previous chapter show that the first seventy six amino acids of H-NS are capable of forming high order oligomeric species in a way comparable to those observed for the wild type protein (Smyth et al., 2000). However, the slightly shorter constructs; H-NS_{1-71} C21S, H-NS_{1-74} C21S, have a reduced ability to form tetramers and therefore a more homogeneous protein sample of these constructs can be prepared increasing the chance of identifying crystallisation conditions for these proteins. Whilst H-NS_{1-71} C21S and H-NS_{1-74} C21S protein constructs are not functional oligomerisation domains the solving of their structure could answer some of the outstanding questions about the structure of H-NS and how the structure dictates the protein's function. Based on the NMR data presented in section 3.6 these two constructs, H-NS_{1-71} C21S and H-NS_{1-74} C21S, are expected to have an intact third helix which corresponds with the full predicted coiled-coil region. The solving of the structure of one of these constructs would demonstrate which of the alternative NMR structures of the H-NS oligomerisation domain represents the correct orientation of the coiled-coil along with the positions of the first and second helices of the oligomerisation domain. This would provide some indication of the relative position of the DNA binding domains when an H-NS homodimer is formed.

The structure of a functional oligomerisation domain, where the predicted fourth helix is intact, would offer a greater potential for understanding how the H-NS protein functions. The structure would allow predictions to be made about how the dimers bind to each other to form tetramers and higher order oligomeric species and how these interactions are influenced by changes in the local environmental conditions, such as temperature changes and changes in ionic strength, that have been shown to affect the oligomeric state of H-NS (Ceschini et al., 2000; Ono et al., 2005). However, due to the ability of these constructs to form a range of different oligomeric species in a reversible concentration dependent manner the crystallisation of constructs of a complete functional oligomerisation domain is expected to be extremely difficult and require careful choice of the solvent conditions. Mass spectrometry analysis of the purified H-

NS_{1-77} C21S and H-NS_{1-80} C21S constructs showed that these proteins are more susceptible to proteolytic degradation than the other H-NS oligomerisation domain constructs (Figure 2.14). There was also some probable proteolytic cleavage at the C-terminal end of the H-NS_{1-74} C21S protein sample observed in the [^{1}H,^{15}N]-HSQC spectra.

Protein samples of H-NS_{1-71} C21S, H-NS_{1-74} C21S and H-NS_{1-83} C21S were expressed and purified and subjected to an iterative process of optimising the crystallisation conditions. Figure 4-6 shows a flow diagram representation of the process of crystal optimisation indicating the various decision points that have been used to guide these experiments. H-NS_{1-77} C21S and H-NS_{1-80} C21S were not prepared for crystallisation trials as they were found to be less resistant to proteolytic degradation than the slightly longer H-NS_{1-83} C21S construct suggesting that the C-terminal end of these proteins is not folded. The [^{1}H,^{15}N]-HSQC of H-NS_{1-73} C21S also shows signs of degradation over a period of one month stored at 4°C, so crystallisation trials of this protein and the shorter H-NS_{1-71} C21S constructs were supplemented with the serine protease inhibitor AEBSF.

4.7 Crystallisation Trials for H-NS Oligomerisation Domain Constructs

The preparation of protein crystals is an iterative process beginning with a broad screening of crystallisation conditions to identify intial crystallisation conditions that can be further optimised to produce crystals large enough to suspend in a cryoloop, so that their diffraction properties can be tested. The crystallisation condition parameters such as the solvent conditions, temperature and crystallisation setup can all be changed, to improve the order of the crystal lattice and therefore the crystal diffraction resolution. Post crystallisation soaks, where the crystal is incubated for a period of time in mother liquor (the condition that produced the protein crystal) supplemented with additional reagents, may also be required for optimal freezing of the crystal, to allow data collection at 100 K, or to dehydrate the crystal to improve the order of the crystal lattice. The process used to prepare and optimise protein crystals of H-NS oligomerisation domain constructs is outlined in Figure 4-6.

Initially, the three constructs chosen for crystallisation trials were tested using a sparse matrix of crystallisation conditions from commercial suppliers. These sparse matrix conditions test a range of pH conditions, precipitant types and concentrations and various salts to sample a broad range of crystallisation space but with a strong bias

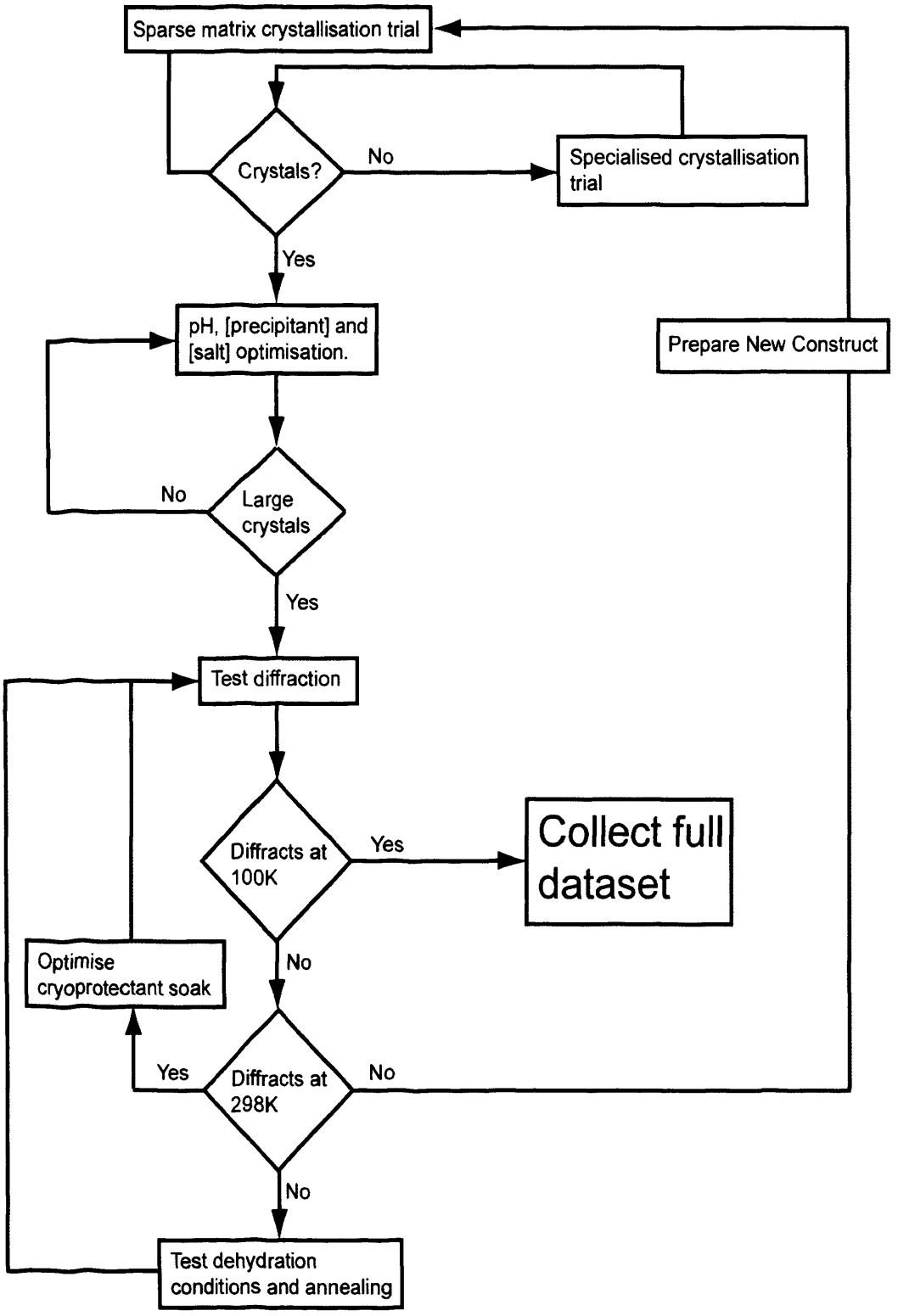


Figure 4-6. A flow diagram showing the iterative process of optimising crystals of the H-NS oligomerisation domain constructs. The rectangular shaped boxes outline the experiments and the diamond shaped boxes show the decision points in the process.

towards conditions previously identified for the crystallisation of other proteins (Jancarik and Kim, 1994).

For each protein the crystallisation trial conditions were tested at either 20°C or 16°C, using the vapour diffusion and microbatch crystallisation methods. At regular time intervals the crystallisation trials were examined using a microscope for the presence of precipitate, phase separation, spherulites, micro crystals or larger crystals.

Small micro crystals were observed in a number of conditions in the crystallisation trial of H-NS_{1-71} C21S. No crystals were identified in the sparse matrix crystallisation trial of H-NS_{1-74} C21S or H-NS_{1-83} C21S. It is possible that crystals would develop at a different temperature or protein concentration for H-NS_{1-74} C21S and H-NS_{1-83} C21S.

However, it could also be the case that these two constructs are more difficult to crystallise than the H-NS_{1-71} C21S construct. HNOE NMR spectroscopy analysis of H-NS_{1-74} C21S showed five unstructured amino acids present in this protein (section 3.6). The presence of this unstructured region would create a region of heterogeneity in the protein that could disrupt the formation of a crystal. Studies using hydrogen deuterium exchange mapping by mass spectrometry have demonstrated that the removal of disordered regions within a protein can significantly improve the likelihood of producing protein crystals of high resolution suitable for structure determination (Pantazatos et al., 2004).

The lack of crystals for the H-NS_{1-83} C21S construct does not appear to be due to the presence of a disordered region. This construct also contains all of the predicted secondary structure elements and highly conserved amino acids within the oligomerisation domain and is therefore expected to fold correctly when over expressed in *E. coli*. However the SEC and AUC studies show that this protein forms a heterogeneous protein solution containing a range of different oligomeric species.

The crystallisation trial was set up with H-NS_{1-83} C21S sample prepared in 10 mM MOPS pH 7.0 as the protein sample is more homogeneous and less likely to form higher order oligomeric species at low ionic strength conditions. However the commercially available sparse matrix screens use ionic strength as one of the parameters for the crystallisation process, along with pH and precipitant. Only 15 of the 96 crystallisation conditions tested do not contain any inorganic salt in the crystallisation condition, so the crystallisation trial will affect the oligomeric state of the H-NS_{1-83} C21S protein.

The crystallisation trial for H-NS_{1-71} C21S identified several conditions that produced small crystals, summarised in Table 4-2, although the crystals observed were not of sufficient size to test their diffraction resolution.

4.8 Optimising the Crystallisation of H-NS_{1-71} C21S

After identifying various crystallisation conditions for H-NS_{1-71} C21S that produced microcrystals the crystallisation conditions needed improving to produce larger crystals. The conditions produced crystals relatively quickly (within two weeks) with hundreds of microcrystals growing in the single drop. The pH, precipitant concentration or salt concentration was changed so that conditions that produce single crystals large enough to pick up with a fibre loop, so that their X-ray diffraction properties could be tested, could be identified. Initially a relatively broad grid screen was used to improve the pH and precipitant concentration in the crystallisation conditions followed by further grid screens where the salt concentration or precipitant concentration was varied.

Large hexagonal crystals were prepared in 100 mM sodium cacodylate pH 6.4, 20-24% (v/v) 2-methyl-2,4-pentanediol, 50 mM magnesium acetate using the hanging drop crystallisation method (Figure 4-7). As an alternative organic solvent, isopropanol was also used as a precipitant and large crystals suitable for X-ray diffraction testing were produced in 100 mM sodium cacodylate, 23% (v/v) isopropanol, 100 mM sodium acetate. These crystals take approximately two weeks to grow but dissolve if left for over a month in the hanging drop setup.

The larger crystals grown in the condition containing 2-methyl-2,4-pentanediol diffracted to 4.6 Å resolution. A complete dataset has not been collected for these crystals due to a vacuum failure at the ESRF synchrotron radiation facility which meant that the crystals could not be tested.

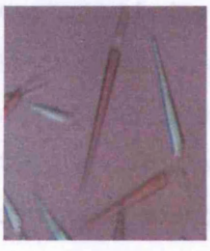
The crystals prepared in the isopropanol containing condition did not diffract as strongly with the larger crystals diffracting to only 8 Å.

Table 4-2. The conditions identified from the sparse matrix crystallisation trial screening of 710 μM H-NS_{1-71} C21S by vapour diffusion. Unless otherwise stated the drop was clear until the appearance of crystals was observed.

Salt	Buffer	Precipitant	Crystal Description	Time taken for Crystals to Appear
None	0.1M sodium citrate pH 5.6	35% tert-butanol	Small whisker micro crystals	5 days
0.2 M magnesium acetate	0.1M sodium cacodylate pH 6.5	30% (v/v) 2-methyl-2,4-pentanediol	Micro crystals	5 days. (Precipitate after 51 days)
None	0.1M HEPES pH 7.5	35% (v/v) 2-methyl-2,4-pentanediol	Small whisker micro crystals	14 days
None	0.1M HEPES pH 7.5	10% (v/v) 2-propanol, 20% PEG 4000	Small whisker micro crystals	51 days (Form from initial precipitate)
None	0.1M Tris pH 8.5	25% tert-butanol	Small whisker micro crystals	51 days (Form from initial precipitate)
None	0.1M Tris pH 8.5	1.0M $(NH_4)_2HPO_4$	Trapezoid shaped crystal	51 days
0.2 M tri-sodium citrate dihydrate	0.1M Na cacodylate pH 6.5	30% (v/v) 2-propanol	Micro crystals	Spherulites after 5 days, micro crystals after 51 days
0.2 M sodium thiocyanate	None	20% (w/v) PEG 3350	Small Crystal plates	51 days. (Precipitate after 5 days. Whiskers appear after 14 days).

Trapezoid shaped crystal
Crystal of H-NS_{1-71} C21S grown in 100 mM Tris pH 8.5, 1.0 M $(NH_4)_2HPO_4$

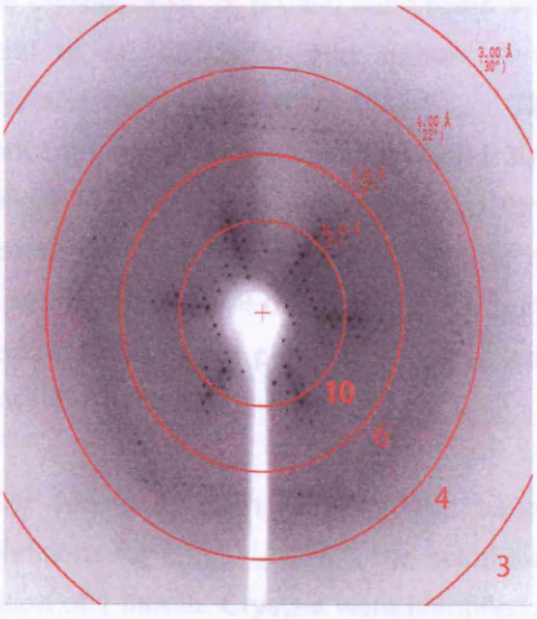
Hexagonal shaped crystal
Crystals of H-NS_{1-71} C21S grown in 100 mM mops pH 6.0, 20% MPD,
Photo taken using a polariser showing the birefringent property of the crystal.



Hexagonal based pyramid shaped crystals
Crystals of H-NS_{1-83} C21S grown in 50 mM mes pH 6.0, 10% MPD, 5mM mops
Photo taken using a polariser showing the birefringent property of the crystals.

Figure 4-7. Protein crystals of H-NS constructs.

A.



B.

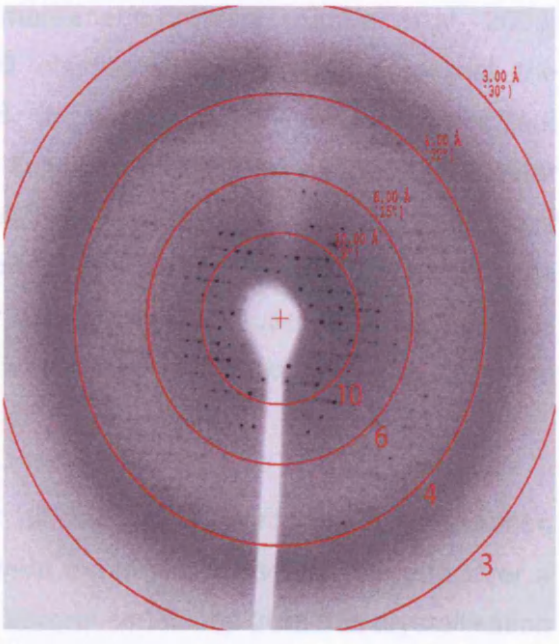


Figure 4-8. A. Diffraction pattern for a crystal of H-NS_{1-71} C21S B. The diffraction pattern collected for a crystal of H-NS_{1-83} C21S. Both diffraction patterns were collected from 0° to 0.5° rotation of the crystal with 15 minute exposure to the X-ray beam using a Rigaku RAXIS IV image plate system. The red circles and numbers show the position the reflections would have for a diffraction resolution of 10 Å, 6 Å, 4 Å and 3 Å respectively.

4.9 PREPARING PROTEIN CRYSTALS OF H-NS_{1-83} C21S

4.9.1 Creating a Crystallisation Screen Designed to Improve the Homogeneity of the Protein Sample

As no crystals were identified using the molecular dimensions sparse matrix crystallisation conditions, an alternative crystallisation screen was needed to produce crystals of H-NS_{1-83} C21S. At low ionic strength H-NS_{1-83} C21S protein is in equilibrium between a homodimer and homotetramer whereas increasing the ionic strength allows higher order oligomeric species to be formed. In a recent review of the structures deposited in the protein databank it was found that the precipitant concentration used to crystallise protein complexes is generally lower than that required to crystallise monomeric proteins. For example, the majority of protein complexes crystallised in 10% to 20% polyethylene glycol (PEG) solutions whereas 20% to 30% PEG is commonly found in the commercially available sparse matrix screens developed for monomeric protein crystallisations (Radaev *et al.*, 2006). PEG is the most commonly used precipitant for the crystallisation of both protein complexes and monomeric proteins whose structures have been deposited into the protein databank. Ammonium sulphate and 2,4-methylpentanediol (MPD) are also commonly used as precipitants for the crystallisation of protein complexes and monomeric proteins. (Kimber et al., 2003; Peat et al., 2005). A grid screen of 190 conditions was prepared to test the crystallisation potential of H-NS_{1-83} C21S in conditions of varying precipitant concentration and pH. 10%, 20% and 30% solutions of MPD and various molecular weight PEG solutions were prepared for each pH condition to test the precipitant concentration range commonly found to crystallise proteins. Every 0.5 pH units was tested between pH 4.0 and 9.0; each buffer was chosen because its pKa value was close to the desired pH . No inorganic salt was added to the conditions to keep the ionic strength relatively low.

After setting up both hanging drop vapour diffusion and microbatch crystallisation conditions for H-NS_{1-83} C21S, the solutions were examined at regular intervals over a period of 1 month. Crystals were identified in several conditions from this crystallisation trial screen, centred on two different pH conditions. First, crystals were observed in MPD and PEG 400 precipitant conditions with 100 mM MES pH 6.0 as the buffer. These conditions produced small crystals that appeared to have a diamond shape morphology. The second group of crystallisation conditions has higher molecular weight PEG as the precipitant and pH conditions around pH 8.5. The crystals formed in

the second set of conditions are described as whiskers. These are clusters of crystals that have nucleated from the sides of other crystals so that small crystals growing in different directions are connected to each other. The crystallisation conditions identified are outlined in Table 4-3 and Table 4-4.

4.9.2 Optimising the Crystallisation of H-NS_{1-83} C21S

As described previously for the optimisation of crystallisation conditions for H-NS_{1-71} C21S, crystallisation trials were prepared to test small changes in the MPD or PEG concentration as well as the pH of the buffer in order to identify improved crystallisation conditions for H-NS_{1-83} C21S.

For the diamond shaped crystals initially grown in 100 mM MES pH 6.0 with either MPD or PEG 400 as the precipitant, the pH and precipitant concentration was varied to reduce the level of nucleation in the crystallisation condition so that single large crystals could be grown. These tests showed that as the pH increases between pH 5.0 and 6.8 the percentage MPD or PEG 400 required to produce crystals also increases. The theoretical pI for the H-NS_{1-83} C21S construct, calculated from its amino acid sequence, is pH 4.79 (Bjellqvist et al., 1993). As the conditions at higher pH require a higher concentration of MPD and PEG 400 to produce crystals these conditions were pursued further not only because larger crystals were produced but also because these conditions would require less cryoprotectant to stabilise the crystals. The addition of cryoprotectant can disrupt the crystal packing of these crystals and large changes in the solvent around the crystal should be avoided whenever possible. The larger crystals (0.5 mm for the longest dimension) grew when a 1 µL solution of 100 mM MES pH 6.2, 20% MPD was added to 1 µL 2 mM H-NS_{1-83} C21S in 10 mM MOPS pH 7.0 and incubated at 16°C under paraffin oil. The larger crystals diffracted to 5-7 Å resolution at both room temperature and at 100 K (after soaking for 1 minute in 55 mM MES, pH 6.2, 12% MPD, 20% glycerol). As the crystals were relatively large the poor resolution of the crystals is probably due to small variations in the packing of the protein molecules within the crystal lattice so that the crystal lattice is not perfectly uniform throughout the crystal.

Table 4-3. Crystallisation conditions identified using microbatch crystallisation setup for H-NS_{1-83} C21S from the low ionic strength grid screen. 1 µL of each condition was mixed with 1 µL 2046 µM H-NS_{1-83} C21S dissolved in 10 mM mops pH 7.0 and incubated at 16°C.

Crystallisation Method	Buffer	Precipitant	Crystal Description	Time taken for Crystals to Appear
Microbatch	0.1 M MES pH 6.0	10% (v/v) MPD	Small diamond crystals	12 days
Microbatch	0.1 M MES pH 6.0	30% (v/v) MPD	Small diamond crystals	2 days
Microbatch	0.1 M MES pH 6.0	10% (v/v) PEG 400	Small diamond crystals	12 days
Microbatch	0.1 M MES pH 6.0	20% (v/v) MPD	Small diamond crystals	12 days
Microbatch	0.1 M MES pH 6.0	10% (w/v) PEG 4000	Small diamond crystals	12 days
Microbatch	0.1 M tris base pH 8.5	30% (w/v) PEG 4000	Whiskers	2 days
Microbatch	0.1 M bicine pH 9.0	30% (w/v) PEG 4000	Whiskers	2 days
Microbatch	0.1 M tris base pH 8.5	20% (w/v) PEG 6000	Whiskers	12 days
Microbatch	0.1 M tricine pH 8.0	30% (w/v) PEG 6000	Whiskers	12 days

Table 4-4. Crystallisation conditions identified using hanging drop crystallisation setup for H-NS_{1-83} C21S from the low ionic strength grid screen. 1 µL of each crystallisation condition was added to 1 µL of 1023 µM H-NS_{1-83} C21S in 10 mM Mops pH 7.0 and suspended above 0.5 mL of the crystallisation trial condition, incubated at 16°C.

Crystallisation Method	Buffer	Precipitant	Crystal Description	Time taken for Crystals to Appear
Hanging drop	0.1 M MES pH 6.0	20% (v/v) MPD	Small diamond crystals	Grew from precipitate after 13 days
Hanging drop	0.1 M pipes pH 6.5	10% (w/v) PEG 400	Whiskers	Grew from precipitate after 25 days
Hanging drop	0.1 M tricine pH 8.0	20% (w/v) PEG 2000	Whiskers	13 days
Hanging drop	0.1 M tris base pH 8.5	20% (w/v) PEG 2000	Whiskers	5 days
Hanging drop	0.1 M bicine pH 9.0	20% (w/v) PEG 2000	Whiskers	13 days

4.10 IMPROVING THE DIFFRACTION RESOLUTION OF THE PROTEIN CRYSTALS

4.10.1 Additive Screen

An additive screen contains a range of different types of chemical reagents that can be added to a crystallisation condition in order to change protein to protein or protein to solvent contacts. Divalent cations have been shown to be required for crystallisation of several proteins by promoting protein to protein contacts within the crystal. The diffraction resolution of crystals of the leucine, isoleucine, valine binding protein (LIVBP) improved from 2.7 Å to 1.7 Å by the addition of 1 mM $CdCl_2$ into the crystallisation condition (Trakhanov and Quiocho, 1995). Seven cadmium binding sites were identified on the surface of the protein with various coordination geometries; four hexameric, two pentameric and one tetrameric coordination site involving histidine, aspartate or glutamate sidechains. Protein stabilising agents such as glycerol and sucrose can stabilise flexible regions of the protein and prevent aggregation. The presence of glycerol or another polyol cosolvent that preferentially associates with polar molecules rather than hydrophobic molecules can stabilise the protein structure, shifting the equilibrium towards the folded state compensating for the destabilising effects of using non-polar solvents such as MPD or PEG as precipitating agents (Sousa, 1995). Detergents have also been successfully used as additives for the crystallisation of proteins where aggregation is a problem (Cudney et al., 1994). For the additive screen of H-NS_{1-83} C21S 0.25 µL of each additive stock solution was added to 1 µL of 100 mM MES pH 6.0, 12% MPD and 1.25 µL 2 mM H-NS_{1-83} C21S in 10 mM MOPS pH 7.0. All of the divalent cations tested caused precipitation of the protein sample. However, crystals were observed in the presence of phenol, dioxane, spermidine and spermine. Phenol is usually considered as a dissociating reagent and may help make the protein sample more homogeneous by disrupting high order oligomerisation interactions of H-NS. Dioxane is a volatile organic solvent which may stimulate nucleation as it evaporates from the under oil conditions, reducing the volume of the drop and effectively increasing the protein and precipitant concentrations. Spermine and spermidine are polyamine chemicals that can act as electrostatic crosslinkers stabilising protein to protein interactions within the crystal (Cudney et al., 1994).

The diffraction properties of the larger crystals from these conditions were tested. Only the crystals grown in the presence of phenol showed any improvement in the diffraction

resolution observed, diffracting to 4 Å resolution compared with approximately 7 Å resolution observed for crystals prepared in the absence of phenol (Figure 4-8B).

A selenomethionine labelled protein sample of H-NS_{1-83} C21S was prepared and crystallised under these conditions. Again the pH and MPD concentration were varied to optimise the crystallisation condition but this time the phenol concentration was also varied. The best crystal was produced in 50 mM MES pH 6.2, 10% (v/v) MPD, 10 mM phenol, showing that the incorporation of selenium atoms into the protein does not affect the preparation of protein crystals. Twenty repeats of this condition were prepared but only three wells yielded any crystals suggesting that this condition is only just sufficient to cause crystal nucleation on a few occasions and is therefore close to being a metastable condition, the condition optimal for crystal growth but where crystals cannot nucleate.

4.10.2 Crystal Dehydration Methods

There are several reports in the literature concerning the use of crystal dehydration to improve the diffraction resolution of a crystal. Dehydration reduces the volume of the unit cell, forcing the protein molecules closer together which, in some circumstances, can result in the formation of additional crystal contacts or an improvement in the uniformity of the crystal packing resulting in a better overall X-ray diffraction resolution. Perhaps the best illustration of dehydration as a means of improving crystal diffraction resolution requires the use of a free mounting system (Kiefersauer et al., 2000; Kiefersauer et al., 1996). A free mounting system allows the user to alter the atmospheric humidity around the crystal in a controlled manner so that the crystal can be gradually dehydrated without the need for changing the chemical composition of the crystal. Experiments using CO dehydrogenase from *Oligotropha carboxidovorans* demonstrated an improvement in the diffraction resolution from 3 Å to 2.2 Å using a 6% reduction in the humidity around the crystal. Gradual reduction of the humidity by a further 5% caused increased mosaicity in the crystal and a further 3% reduction in humidity lead to damage to the crystal packing and a complete loss of diffraction resolution. Rehydration of the crystal using the conditions that gave a 6% reduction in humidity was able to restore the crystal packing and 1.8 Å resolution data could be collected. This approach was found to be more successful with crystals grown in concentrated salt solutions than with crystals where PEG or MPD had been used as the precipitant (Kiefersauer *et al.*, 2000). In the absence of a free mounting system an alternative method is needed to dehydrate the crystal. Chemical soaks with various

drying conditions have been successfully used to improve the diffraction resolution of protein crystals (Heras et al., 2003; Heras and Martin, 2005; Petock et al., 2001; Hare et al., 2006).

Initial testing of the crystal soak conditions was performed by transferring five to 10 small crystals of H-NS_{1-83} C21S into a 5 µL stabilising solution of 50 mM MES pH 6.2, 12% MPD, 10 mM phenol supplemented with various concentrations of PEG reagents as described in Table 4-5. The crystals were then monitored for stability under these conditions. Observations of the effect of the soak conditions on the crystal morphology and its birefringence properties were recorded. Birefringence is a measure of the order of the crystal lattice. When a crystal is placed between two polarising filters it will have a different colour to the background as shown in Figure 4-7. A loss of birefringence indicates a disruption to the regular ordering of the crystal packing. The crystals appear to tolerate slightly higher MPD concentrations but an increase in MPD from 10% to 20% (v/v) results in a loss of birefringence. Crystals of H-NS_{1-83} C21S were soaked in a stabilising solution containing an additional 3% MPD (15% in total) were left for either 30 minutes or overnight at 19°C. After incubation the crystals were transferred for 1 minute into a 2 µL drop containing this solution supplemented with 20% (w/v) glycerol to act as a cryoprotectant. The crystals were then flash frozen at 100K and their diffraction properties measured using a 15 minute exposure time. These crystals diffracted to 7-8 Å showing no improvement in the diffraction resolution.

The crystal soaks when the stabilising solution was supplemented with various PEG solutions showed varying results, depending on the molecular weight of the PEG used. The crystals did not dissolve under any of the conditions tested but the crystal soaks in 25% (w/v) PEG 2000 or 5% to 15% (w/v) PEG 6000 supplemented solutions showed a loss of birefringence suggesting that these precipitants are not suitable for dehydrating the H-NS_{1-83} C21S crystals. A loss of birefringence indicates a disruption to the order of the crystal packing. The diffraction properties of crystals soaked in stabilising solution supplemented with 5% PEG 400 for 5 minutes was tested. The two crystals tested, after incubation at room temperature in this condition, showed a loss of diffraction resolution with no observable reflections, after a 15 minute exposure to X-rays.

When testing dehydration conditions the most promising conditions should cause the crystals to crack if the precipitant concentration is too high. Cracking of the crystal indicates disruption to the planes in the crystal lattice brought about by the loss of solvent from the crystal. Only the PEG 8000 supplemented crystal soaks showed any

Table 4-5. Visual analysis of the effect of various crystal soak solutions on the birefringence and morphology of crystals of H-NS_{1-83} C21S.

Crystal Soak	Observation	Crystal Soak	Observation
3% (v/v) MPD	Crystal is stable for 1 hour. Birefringence is maintained.	5% (w/v) PEG 4000	Crystals appear stable in this condition for 1 hour. Birefringence maintained
5.5% (v/v) MPD	Crystal is stable for 1 hour. Birefringence is maintained.	10% (w/v) PEG 4000	Crystals appear stable in this condition for 1 hour. Birefringence maintained
8% (v/v) MPD	Crystal loses birefringence after 5 minute soak. Crystal does not dissolve.	15% (w/v) PEG 4000	Crystals appear stable in this condition for 1 hour. Birefringence maintained
5% (v/v) PEG 400	Crystals appear stable in this solution for up to 30 minutes. Birefringence maintained.	25% (w/v) PEG 4000	Crystals appear stable in this condition for 1 hour. Birefringence maintained
10% (v/v) PEG 400	Birefringence maintained for 30 minutes but edges of the crystals are less well defined.	5% (w/v) PEG 6000	Crystals do not dissolve in this condition but birefringence is lost after 5 minute soak.
15% (v/v) PEG 400	Birefringence maintained for 30 minutes but edges of the crystals are less well defined.	10% (w/v) PEG 6000	Crystals do not dissolve in this condition but birefringence is lost after 5 minute soak.
20% (v/v) PEG 400	Birefringence maintained for 75 minutes.	15% (w/v) PEG 6000	Crystals do not dissolve in this condition but birefringence is lost after 5 minute soak.
45% (v/v) PEG 400	Loss of birefringence after 20 minutes.	5% (w/v) PEG 8000	Crystals are stable for 30 minutes. Birefringence maintained.
5% (w/v) PEG 2000	Crystals appear stable in this condition for up to 45 minutes. Birefringence maintained	10% (w/v) PEG 8000	Crystals show cracking at the surface but birefringence maintained after 55 minute soak.
10% (w/v) PEG 2000	Crystals appear stable in this condition for up to 1 hour. Birefringence maintained	15% (w/v) PEG 8000	Crystals show cracking at the surface. Loss of birefringence observed after 55 minute soak.
15% (w/v) PEG 2000	Crystals appear stable in this condition for up to 1 hour 45 minutes. Birefringence maintained.		
25% (w/v) PEG 2000	Crystals appear stable in this condition for 5 minutes but loss of birefringence observed after 1 hour.		

visible cracking on the surface of the crystals tested. 15% (w/v) PEG 8000 caused cracking of the surface of the crystal and a loss of birefringence but the lower concentrations of PEG 8000 (5% and 10%) showed no loss of birefringence after 30 minutes and 55 minutes respectively. The condition containing 10% (w/v) PEG 8000 showed some cracking on the surface of the crystals suggesting that dehydration was occurring but that this condition contained slightly too high a concentration of PEG 8000 to improve the diffraction resolution. Crystals suitable for testing their X-ray diffraction were soaked for 30 minutes in 5%, 6%, 7%, 8% and 9% PEG 8000 solutions. The crystals showed no improvement in the diffraction resolution, diffracting to between 5 and 7 Å resolution, after a 15 minute exposure to the X-ray beam.

4.10.3 Crystal Cross-Linking

To improve the robustness of crystals for dehydration and cryoprotection soaks protein crystals can be cross-linked using glutaraldehyde soaks to create additional protein to protein contacts that stabilise the crystal packing (Lusty, 1999; Quiocho and Richards, 1964). 0.125% (v/v) glutaraldehyde solution in 50 mM MES pH 6.0, 11% MPD, 5 mM Mops pH 7.0, 25% glycerol, 10 mM phenol was prepared. Crystals of H-NS_{1-83} C21S were incubated in 2 µL drops of this solution for 1 minute, 5 minutes and 9 minutes before flash freezing the crystal. The diffraction properties of the crystals were then tested. The crystals soaked for 9 minutes appeared damaged by observation under the microscope and showed a loss of birefringence. The crystals soaked for 5 minutes in the glutaraldehyde solution maintained their morphology but no diffraction spots were observed after a 15 minute exposure to the X-ray beam. The crystals soaked for 1 minute only diffracted to 10 Å resolution. Glutaraldehyde cross linking under these conditions damages the crystal packing of the H-NS_{1-83} C21S crystals.

4.10.4 Crystallisation in an Agarose Gel.

Agarose gels have previously been used to promote the growth of high quality protein crystals (Lorber et al., 1999; Sica et al., 1994; Vidal et al., 1999). Gellified media is characterised by the absence of convection currents and the prevention of crystal sedimentation, mimicking the effects of microgravity. These conditions may favour higher internal order and generate crystals with lower mosaicity and greater diffraction resolution, as demonstrated with crystals of thaumatin (Lorber et al., 1999) and lysozyme (Vidal et al., 1999). 2% (w/v) low melting point agarose (sigma) was melted at 65°C and then allowed to cool to 40°C using a heating block. The molten agarose

was then mixed with the crystallisation condition (100 mM MES pH 6.2, 20% (v/v) MPD, 20 mM phenol) to give a final concentration of 0.3% agarose. 1 µL of this solution was immediately added to 1 µL of 2 mM H-NS_{1-83} C21S, in 10 mM MOPS pH 7.0, dispensed under paraffin oil in a terasaki plate. Five protein crystals, prepared using this method, were mounted on a fibre loop and their X-ray diffraction pattern collected. No improvement in the diffraction resolution was observed compared to crystals grown in the same conditions in the absence of the agarose gel.

4.11 X-RAY DIFFRACTION

The best crystals of H-NS_{1-71} C21S and H-NS_{1-83} C21S diffracted to about 4 Á resolution (Figure 4-8). A full data set was collected for a crystal of H-NS_{1-83} C21S in 1° increments for 65° of rotation of the crystal. Indexing of the reflections using the Crystal clear 1.3.6 SP3.3 software (Rigaku MSC) suggests that the crystal packing corresponds to a trigonal or hexagonal space group with unit cell dimensions of A = 130.5 Á, B= 130.5 Á and C = 55.0 Á. The unit cell angles are α = 90°, β = 90° and γ = 120° with the most likely space group being one either $P6_1$ or $P6_5$. The low intensity of the reflections and the large volume of the unit cell indicate that the crystals are protein crystals. A crystal formed from one of the solvent components would give a diffraction pattern with larger spacing between the reflections, indicating smaller unit cell dimensions, and stronger intensity reflections.

4.12 MOLECULAR MODELLING

Molecular models of the structure of the H-NS oligomerisation domain were prepared using the two available NMR structures of H-NS to try to create an improved model structure to interpret the X-ray diffraction data by molecular replacement. The parallel coiled-coil structure of H-NS_{2-58} C21S was extended to the predicted end of the third helix at aspartate 68 by comparative modelling with known parallel coiled-coil structures. The antiparallel coiled-coil structure of H-NS_{1-46} was extended, by Dr M. Williams, to residue 68 by extending the third helix using main chain dihedral angle restraints for an α-helix and energy minimisation.

Comparative modelling is based on the observation that proteins with similar sequences adopt similar three dimensional structures (Chothia and Lesk, 1986). To generate a comparative model known protein structures that are homologous to the

sequence being modelled are initially superposed as rigid bodies, using multiple least squares fitting. The sequence of the target protein is then aligned with this consensus sequence, to determine spatial restraints for the modelling process. There are numerous spatial restraints that are used to determine the comparative model structure. Some of the spatial restraints, such as the distances between different secondary structure elements and the relative angles of these regions of secondary structure, are determined from the alignment with the known structures. Other spatial restraints such as Cα-Cα distances, bond lengths, main-chain dihedral angles or side-chain dihedral angles are derived from spatial restraints taken from databases of known protein structures (Sali and Blundell, 1993). In the case of the Modeller software package (Eswar et al., 2003; Sali and Overington, 1994; Sali et al., 1995) used here a database of 105 protein structures is used to determine distributions for these spatial restraints, to create probability density functions for each restraint; these conditional restraints are then implemented by the software when the protein model is being generated (Sali and Overington, 1994). Models are initially generated that satisfy short range restraints, before applying longer range restraints to determine the tertiary structure of the protein. Molecular dynamics with simulated annealing is used to refine the model in an automated way, using the modeller software.

Using the comparative modelling approach, the structure of H-NS_{2-58} C21S was extended to include the amino acids up to Asp68 that is predicted, using secondary structure prediction algorithms, to be the end of the third helix of H-NS. The third helix is predicted to form a coiled-coil structure so the structures of known parallel coiled-coils were used as the template for this comparative modelling process. Initially a consensus coiled-coil template structure was created by superposing the structures of the Sir4 protein coiled-coil domain (PBD file 1PL5) with a coiled-coil of tropomyosin (PDB file 1IC2). The amino acid sequence for the predicted third helix of H-NS, residues T22 to D68, was then aligned with this consensus structure, ensuring that the heptad repeat of the coiled-coils were aligned correctly. A parallel coiled-coil model was generated that best satisfied the spatial restraints from the template. The model structure was then aligned with the residues S2 to R41 from the NMR structure of H-NS_{2-58} C21S (PDB file 1LR1) (Esposito et al., 2002). Residues 41 to 58 were not included in this alignment because the [^{1}H,^{15}N]-HSQC spectra for H-NS_{2-58} C21S and H-NS_{1-74} C21S shows differences in the backbone amide crosspeak positions for amino acids 42 to 58. The new structure alignment was used to generate the final comparative model for H-NS_{1-68} C21S shown in Figure 4-9.

4.12.1 Parallel coiled-coil H-NS_{1-68} C21S Model Structure

The accuracy of a model protein structure, built by comparative modelling, is dependent on the sequence similarity of the structures used to model the protein and that of the target protein. As there are no coiled-coil structures currently available within the protein databank that have greater than 30% sequence identity with H-NS and correspond to the C-terminal end of the oligomerisation domain (residues 40 to 83), any model structure cannot be treated as an accurate representation of the real structure. However, the extended coiled-coil structure generated using Modeller, provides a model for how an intact coiled-coil might dimerise. In the model of H-NS_{1-68} C21S, based on a parallel coiled-coil, the full length of the third helix is involved in forming the coiled-coil interaction, as predicted by the Marcoil coiled-coil prediction algorithm (see Figure 1.7). This coiled-coil prediction is consistent for H-NS and StpA proteins from a number of different enteric bacteria. The model structure was submitted to Procheck (Laskowski *et al.*, 1993) to test the acceptability of the bond angles and bond lengths within the model structure. Serine 21 amino acids from both protomer chains are located in an acceptable region but not the most favoured region of the Ramachandran plot. Asn10 is located in an unacceptable region for both protomer chains. All of the residues within the coiled-coil region of the protein are found in favoured regions of the Ramachandran plot.

Three charged amino acids are found within the heptad repeat of the coiled-coil, Arg 40, Glu 44 and Arg 54, and this protein structure model provides a hypothesis for how these charges are accommodated. The sidechain of Arg 40 points towards the sidechain of Glu 44 from the other protomer chain, potentially forming a stabilising salt bridge. A salt bridge can stabilise charged residues located within coiled-coil proteins. In the structure of the Bcr-Abl oligomerisation domain (pdb file 1k1f), Glu 51, which is located within the heptad repeat, forms a salt bridge with arginine 55 to prevent the negatively charged glutamate disrupting the coiled-coil (Zhao *et al.*, 2002). A similar situation is found in the coiled-coil domain of human EB1 protein (pdb file 1yiB) where Glu 211 and Arg 214 are accommodated within the hydrophobic core of the coiled-coil by forming a salt bridge with each other (Slep *et al.*, 2005). However no salt bridge was identified between amino acids R40 and E44 in the parallel coiled-coil structure of H-NS_{2-58} C21S (Esposito et al., 2002).

The Arg54 sidechain in the model of H-NS_{1-68} C21S points out into the solvent and would therefore not be disruptive to the formation of the coiled-coil. CD has been used

to show that mutation of Arg54 to a cysteine, an amino acid with a shorter polar sidechain, disrupts the folding of H-NS (Schroder et al., 2001).

4.12.2 Anti-parallel coiled-coil H-NS_{1-68} Model Structure

An alternative model of the first 68 amino acids of H-NS was prepared by extending the H-NS_{1-46} NMR structure (PDB file 1ni8) (Bloch et al., 2003). To extend this structure an assumption was made that the third helix of each H-NS protomer extends to residue aspartate 68 following the same trajectory as residues 22 to 46 that are present in the NMR structure (figure 4-10). This model was built by Dr M. Williams using phi and psi backbone torsion angle restraints for an α-helix. This model does not explain why residues Glu44, Ser45 and Ala46 experience a different chemical environment in the H-NS_{1-74} C21S construct as opposed to the H-NS_{1-58} C21S construct or the disruption to the protein folding caused by the R54C mutation discussed in section 4.12.1. It is possible that if H-NS forms an antiparallel coiled-coil, as found in the published NMR structure of *E. coli* H-NS_{2-47}, then the third α-helix would fold back, either upon itself or onto the third helix of the other protomer. However, in the absence of a close homologue structure in the protein databank comparative modelling techniques cannot be used to produce an accurate model for this putative protein fold.

4.13 Discussion

The structure of the oligomerisation domain of H-NS provides a significant challenge for a structural biologist. The oligomerisation domain is able to self-associate to form a homodimer, homotetramer and larger oligomeric species in a reversible concentration dependent manner. To solve a protein structure it is necessary to isolate one oligomeric form of the protein as dynamic exchange between different oligomeric states will reduce the signal intensity in an NMR spectroscopy experiment or the likelihood of forming crystals for X-ray crystallography.

The heterogeneous nature of the H-NS_{1-83} C21S protein sample which may have contributed to the lack of protein crystals produced using commercially available crystallisation screens was overcome by the preparation of a dedicated grid screen that limited the ionic strength of the solvent, in order to minimise the heterogeneity of the protein sample. Using this approach crystals of both unlabelled and selenomethionine labelled H-NS_{1-83} C21S were produced. The crystallisation conditions were optimised by variation of the pH , precipitant concentration, and additive conditions to improve the

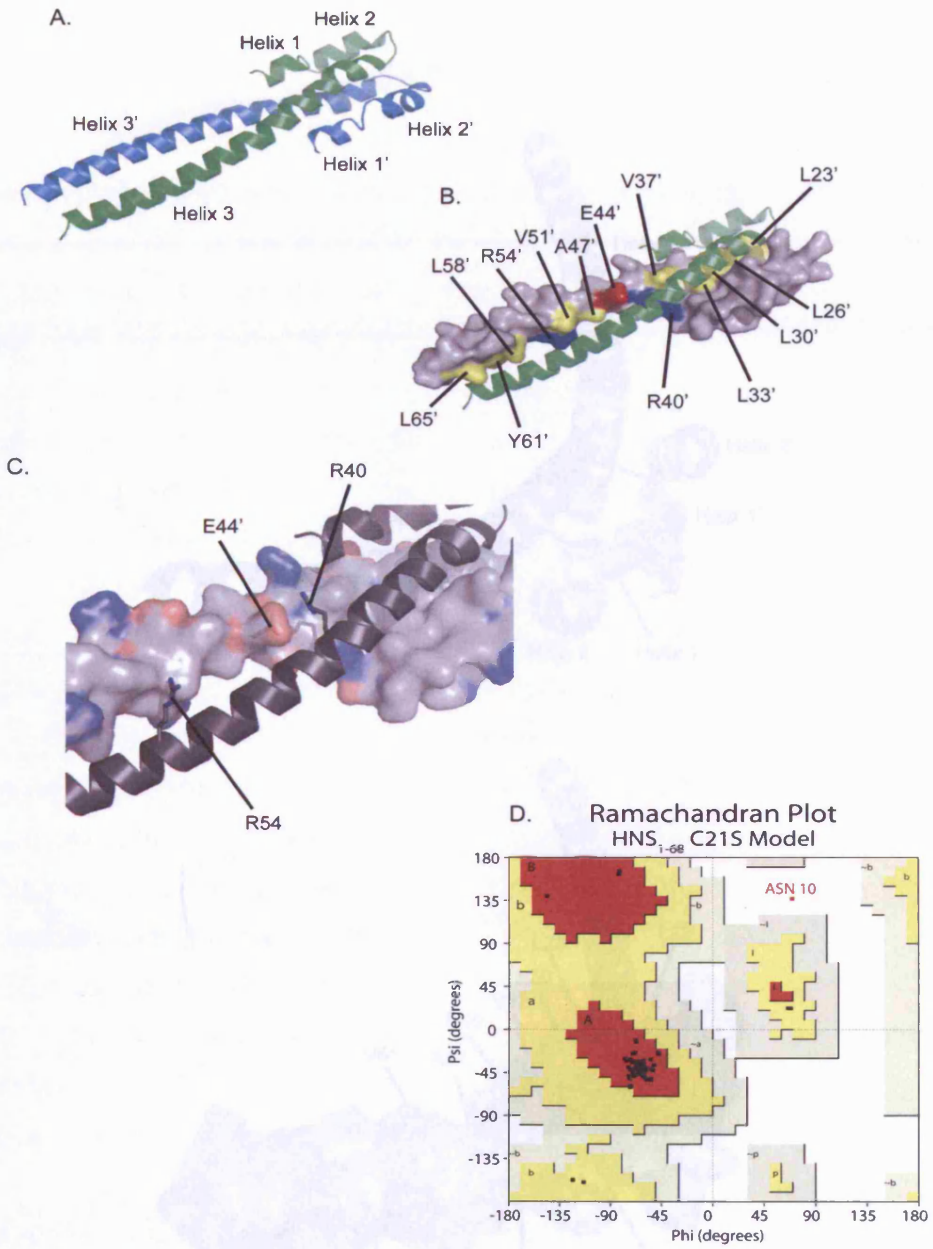


Figure 4-9. Model of an extended parallel coiled-coil structure of H-NS_{1-68} C21S. A. Cartoon depiction of the H-NS_{1-68} C21S homodimer, showing the relative positions of the three α-helices in each protomer. B. A van der Waals surface representation of one of the protomers. Amino acids that comprise the heptad repeat of the coiled-coil are coloured yellow, red or blue for hydrophobic, negatively charged or positively charged amino acids respectively. The other protomer chain is depicted with a cartoon representation. C. A van der Waals surface representation of one of the protomer chains. Charged amino acids are coloured so that oxygen atoms are coloured red and nitrogen atoms coloured blue. The other protomer chain is depicted as a cartoon with selected charged side-chains shown as stick representations. D. A ramachandran plot showing the backbone torsion angles determined for the protein model of H-NS_{1-68} C21S. Strongly favoured regions, allowed regions, generously allowed regions and unacceptable regions are coloured red and yellow, cream and white respectively.

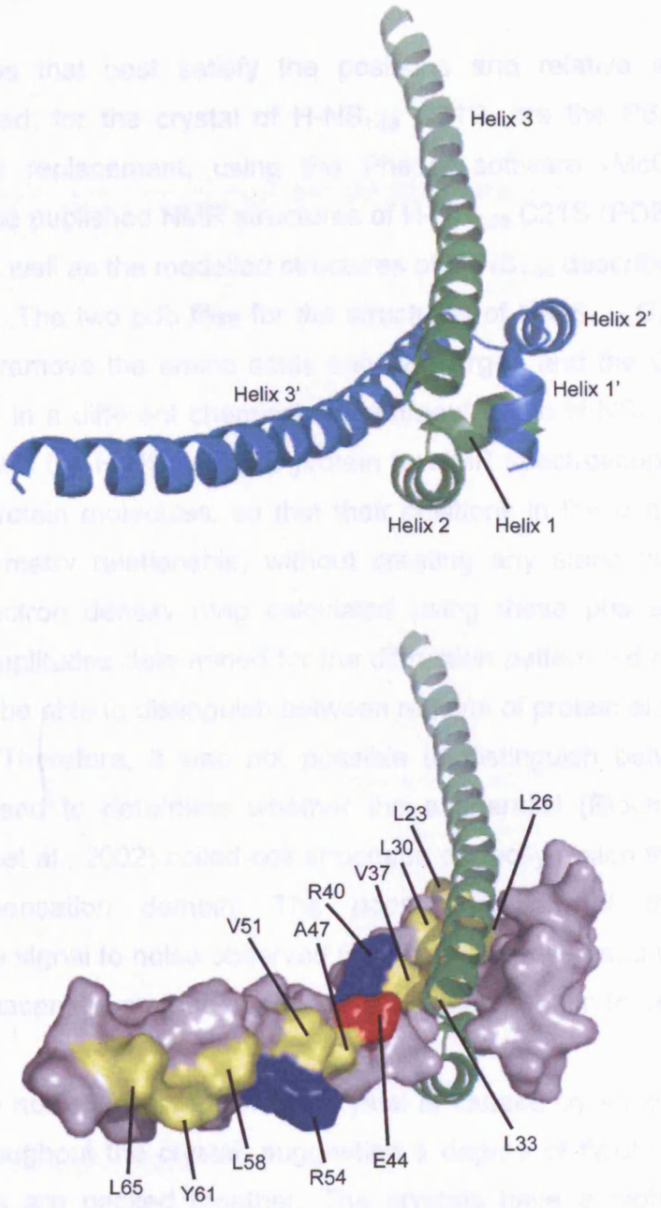


Figure 4-10. Model of H-NS$_{1-68}$ based on the structure of *E. coli* H-NS$_{2-47}$. A. A cartoon depiction of H-NS$_{1-68}$ antiparallel coiled-coil model. B. A van der Waals surface representation of one of the protomer chains is shown with the hydrophobic residues predicted to form the coiled-coil shown in yellow. Arg 40 and Arg 54 (blue) and Glu 44 (red) are highlighted.

diffraction limit of the protein crystals to 4 Å resolution. These crystals are not of sufficiently high enough resolution to enable the structure of the protein to be determined but they may provide a suitable starting point for further optimisation of the crystallisation conditions.

The space groups that best satisfy the positions and relative intensities of the reflections observed, for the crystal of H-NS_{1-83} C21S, are the $P6_1$ and $P6_5$ space groups. Molecular replacement, using the Phaser software (McCoy, 2007), was attempted using the published NMR structures of H-NS_{2-58} C21S (PDB code 1LR1) and H-NS_{2-47} (1NI8) as well as the modelled structures of H-NS_{1-68} described in section 4.15 as search models. The two pdb files for the structures of H-NS_{1-58} C21S and H-NS_{2-47} were modified to remove the amino acids between Arg41 and the C-terminus, which were shown to be in a different chemical environment in the H-NS_{1-73} C21S construct when compared with the H-NS_{1-58} C21S protein by NMR spectroscopy. It was possible to orientate the protein molecules, so that their positions in the unit cell satisfied the space group symmetry relationship, without creating any steric hindrance clashes. However, the electron density map calculated using these phase angles and the structure factor amplitudes determined for the diffraction pattern did not have sufficient signal intensity to be able to distinguish between regions of protein electron density and the background. Therefore, it was not possible to distinguish between the various search models used to determine whether the antiparallel (Bloch et al., 2003) or parallel (Esposito et al., 2002) coiled-coil structures correctly match the structure of the complete oligomerisation domain. The poor resolution of the crystals and corresponding low signal to noise observed for the reflections needs to be improved for the molecular replacement method of phase angle determination to be successful.

The low signal to noise observed for the crystal is caused by small variations in the crystal lattice throughout the crystal, suggesting a degree of flexibility in the way the protein molecules are packed together. The crystals have a high solvent content, reflected in the estimated mosaicity value of 2.0° for the best crystal of H-NS_{1-83} C21S identified. A crystal with a high solvent content will have greater flexibility in the way it is packed, due to large areas of the unit cell being occupied by solvent channels rather than protein molecules. Dehydration of protein crystals has been shown to improve the crystal diffraction resolution and mosaicity, although this can take between a few minutes (Heras and Martin, 2005) and several months (Kuo *et al.*, 2003) to be effective. It may be the case that an improved dehydration method that more slowly dehydrates the crystals would improve the resolution of the H-NS_{1-83} C21S crystals, allowing the structure of a complete oligomerisation domain to be determined.

5 COMPARING THE DNA BINDING PROPERTIES OF H-NS AND STPA

5.1 INTRODUCTION

E. coli H-NS and StpA share 58% sequence identity overall with 68% sequence identity within the DNA binding domain (residues 90 to 134). Despite the close homology between H-NS and StpA, StpA has been shown to have a slightly higher affinity for DNA than H-NS. The binding affinities of H-NS and StpA for a 403 base pairs DNA fragment were shown, by filter retention assay, to be 2.8 μM and 0.70 μM respectively (Sonnenfield et al., 2001). This difference in affinity could be the result of changes between H-NS and StpA within the nucleic acid binding domain directly affecting the interaction with the DNA. Alternatively, differences between the oligomerisation domains of the H-NS and StpA proteins affecting the self-association interactions required to form homodimers and larger oligomeric species could be responsible for this difference in DNA binding affinity. The DNA binding properties of the isolated nucleic acid binding domains from *S. typhimurium* H-NS and *E. coli* StpA are examined using NMR titration experiments and DNA gel shift assays to compare directly the affinity of the two domains for the same DNA fragment and to identify the amino acid residues within the nucleic acid binding domain that are involved in the binding interface. Additional experiments with full length H-NS and StpA have been used to identify the minimum DNA duplex length required for a H-NS homodimer to bind to the DNA.

5.2 TESTING THE INTERACTION OF FULL LENGTH H-NS AND STPA WITH DIFFERENT LENGTHS OF DOUBLE STRANDED DNA

H-NS is thought to preferentially bind to AT rich sequences *in vivo* (Lucchini et al., 2006), although no consensus DNA sequence has been found for this protein (Robison *et al.*, 1998). Double stranded AT-rich DNA fragments were produced by annealing varying lengths of single stranded DNA, from 10 bases to 20 bases, to their complementary DNA molecules at a 1:1 ratio. AT-rich DNA molecules of different lengths, from 10 bases to 20 bases were annealed to their complementary DNA molecule at a 1:1 ratio by heating the solutions in a water bath to 90°C before allowing the solutions to cool slowly to room temperature. Table 5-1 shows the DNA sequence

Table 5-1. Calculated molar extinction coefficients and calculated melting temperatures for oligonucleotides of different length based around a repeating AAT sequence. The molar extinction coefficients were calculated from nucleotide composition (Puglisi and Tinoco, Jr., 1989).

DNA Oligonucleotide	DNA Sequence	Molar Extinction Coefficient/ M^{-1} cm^{-1}	Tm/°C
AAT10mer Forward	5'-GCAATAATCG-3'	96.26	28.0
AAT10mer Reverse	5'-CGATTATTGC-3'	104.66	
AAT12mer Forward	5'-GCAATAATAACG-3'	114.24	32.0
AAT12mer Reverse	5'-CGTTATTATTGC-3'	129.88	
AAT14mer Forward	5'-GCAATAATAATACG-3'	136.6	36.0
AAT14mer Reverse	5'-CGTATTATTATTGC-3'	152.24	
AAT16mer Forward	5'-GCAATAATAATAATCG-3'	158.02	38.9
AAT16mer Reverse	5'-CGATTATTATTATTGC-3'	173.3	
AAT20mer Forward	5'-GCAATAATAATAATAATACG-3'	198.36	45.0
AAT20mer Reverse	5'-CGTATTATTATTATTATTGC-3'	220.88	

of the oligonucleotide primers used to prepare the various lengths of double stranded DNA (10, 12, 14, 16 and 20 base pairs DNA). 2μM DNA duplexes were then incubated in the presence of 10 μM of either purified full length *S. typhimurium* H-NS C21S (H-NS_{FL} C21S) or *E. coli* StpA ($StpA_{FL}$) protein, in 10 mM sodium phosphate, pH 7.0, 100 mM NaCl and 10 μM NaN_3, before analysing the samples by DNA gel shift assay, using an 8% acrylamide gel. DNA samples at the same concentration (2 μM) were also loaded onto the gel as a negative control so as to determine the position of the DNA band in the gel in the absence of protein. O'GeneRuler™ ultra low range DNA size markers (Fermentas) was also loaded on to the gels to enable a comparison between different gels for the size of the gel shift observed when either H-NS_{FL} C21S or $StpA_{FL}$ binds to the DNA.

Figure 5-1 shows the DNA gel shift assays for the binding of H-NS and StpA to varying lengths of duplex DNA. For the 10 base pair DNA molecule, no interaction of the DNA with either H-NS_{FL} C21S or $StpA_{FL}$ was observed as the only band present is at the same level in the gel as the DNA in the 10 base pairs DNA control lane, where no protein was added. When H-NS_{FL} C21S is present with the 12 base pairs DNA duplex, a smeared region labelled B1 is visible in the gel and the size of the band corresponding to free DNA is smaller than in the no H-NS_{FL} C21S control lane, indicating that an interaction between the protein and the DNA does occur. However, a well defined band (B1) is observed for the interaction of $StpA_{FL}$ with the 12 base pairs DNA duplex suggesting that the interaction of $StpA_{FL}$ forms a stronger interaction with the 12 base pairs DNA duplex compared with H-NS. When the 14 base pairs DNA duplex is present a well defined band (B1) is observed for the binding of both H-NS and StpA to this DNA duplex.

The results show that 12 base pairs of DNA duplex is the minimum DNA length required for H-NS_{FL} C21S or $StpA_{FL}$ to bind but in the case of H-NS a more stable interaction is formed when the 14 base pairs DNA duplex is present. Given that the apparent dissociation constant for a monomeric *E. coli* H-NS nucleic acid binding domain with a 127 base pairs DNA duplex is 1 mM, when measured by DNA gel shift assay (Shindo et al., 1995), it is not expected that binding would be observed, with only 10 μM protein, if only one nucleic acid binding domain is accommodated on the DNA. The binding to the 12 base pairs DNA duplex can be explained if there is a cooperative interaction with two nucleic acid binding domains from either H-NS or StpA binding to each DNA molecule. DNA footprinting studies have demonstrated that H-NS binds in the major groove of DNA (Tippner et al., 1994). A B-DNA molecule has 10 base pairs

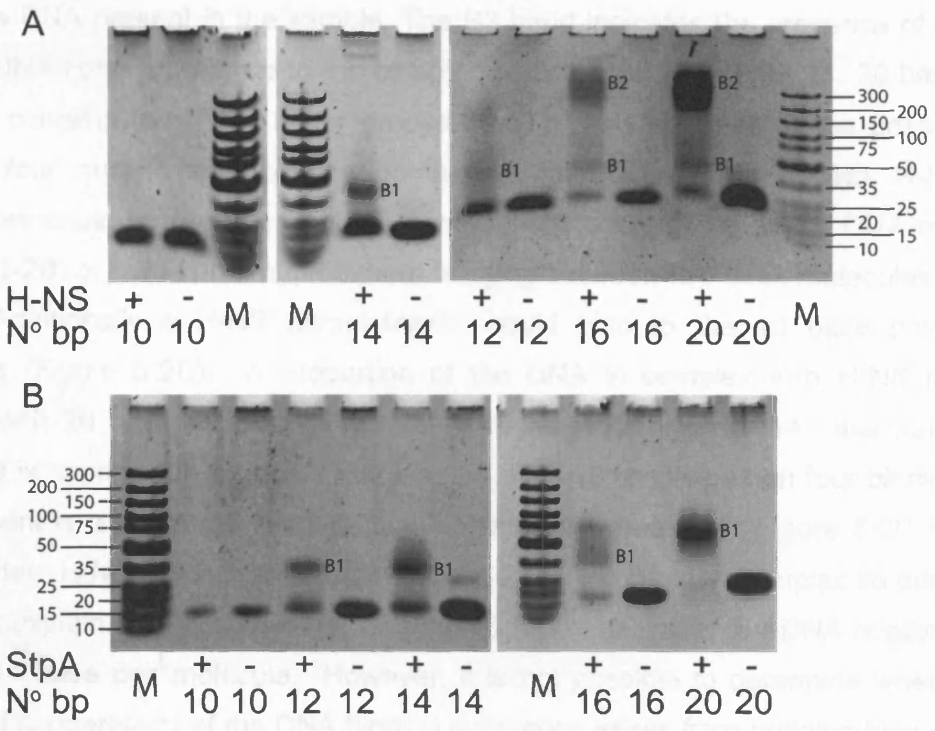


Figure 5-1. DNA gel shift assays testing the binding of 10 μM *S. typhimurium* H-NS$_{FL}$ C21S (A) or *E. coli* StpA$_{FL}$ (B) to different lengths of double stranded DNA. Each length of DNA was run on an 8% (v/v) acrylamide gel in the presence (+) and absence (-) of either H-NS or StpA. Lane M contains O'Gene Ruler Ultra Low DNA size markers. The size of the DNA size marker bands (in base pairs) is shown at the side of the gels. B1 and B2 indicate the two H-NS:DNA complexes observed in the DNA gel shift assay.

per turn. Therefore a 12 base pair DNA duplex will have two equivalent major groove binding sites on opposite 'sides' of the DNA, if viewed from a fixed orientation.

In the case of H-NS when the length of the DNA is extended further to 16 base pairs or 20 base pairs another band on the gel is visible, B2 (Figure 5-1). The B2 band represents a much greater gel shift than that observed when 14 base pairs DNA is present and the proportion of the total DNA present in the B2 band accounts for almost all of the DNA present in the sample. The B2 band indicates the presence of a larger protein:DNA complex relative to the complex corresponding to band B1. 20 base pairs of DNA contains four DNA major groove binding sites so there is the possibility of binding four nucleic acid binding domains. With 20 base pairs DNA H-NS:DNA complexes could form with two H-NS homodimers binding to a single DNA molecule (Figure 5-2B) or two H-NS homodimers bridging between two DNA molecules (Figure 5-2C). Additionally a H-NS homotetramer could bind to the 20 base pairs DNA molecule (Figure 5-2C). A proportion of the DNA in complex with H-NS is much greater with 20 base pairs fo DNA relative to when the shorter 14 base pairs DNA molecule is used. This suggests that binding of H-NS to DNA when four binding sites are present is a cooperative interaction. The model present in Figure 5-2B has two independent H-NS homodimers binding to the 20 base pairs DNA duplex so this model does not explain the increased affinity of H-NS for the 20 base pair DNA relative to the shorter 14 base pair molecule. However, it is not possible to determine whether the observed cooperativity of the DNA binding interaction arises from bridging between two 20 base pair DNA molecules by two H-NS homodimers or from the formation of an H-NS homotetramer.

When StpA is incubated with either the 16 base pairs or 20 base pairs DNA only one band is observed in the gel, corresponding to a DNA:protein complex, B1. When StpA is bound to the 20 base pairs DNA, the size of the DNA band, B1, is not as intense as the band for DNA in absence of StpA. It is possible that a larger StpA DNA complex is formed which does not migrate into the gel, although no dark band is observed at the bottom of the well.

5.3 TESTING THE AFFINITY OF H-NS AND STPA NUCLEIC ACID BINDING DOMAINS FOR DNA

The binding affinities of the nucleic acid binding domains of H-NS and StpA, $H\text{-}NS_{90\text{-}137}$ and $StpA_{91\text{-}134}$, to a AAT20mer double stranded DNA molecule (5'-GCAATAATAATAATACG-3')

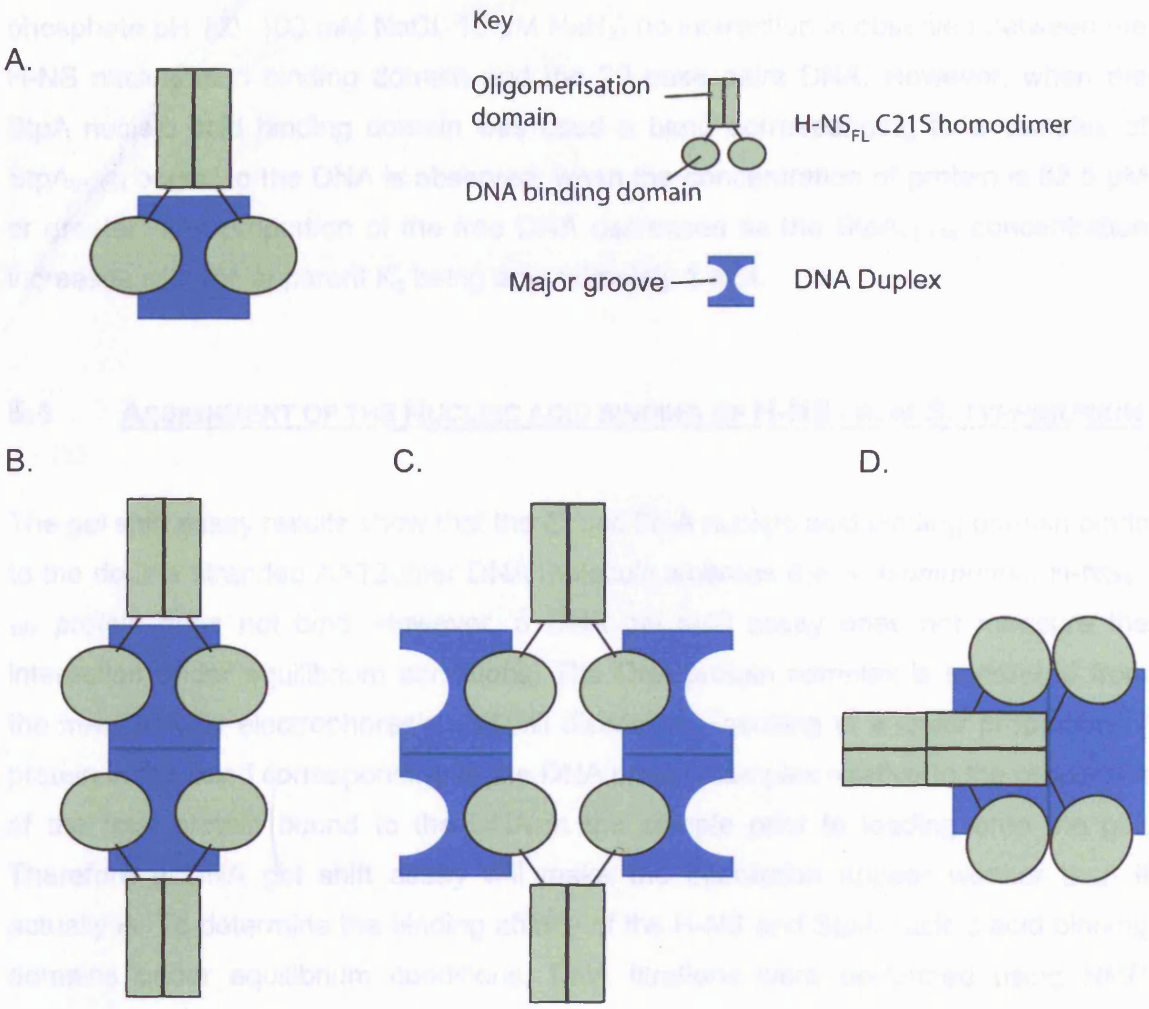


Figure 5-2. Models for the binding of $H\text{-}NS_{FL}$ C21S to DNA. A. The binding of a $H\text{-}NS_{FL}$ C21S homodimer to a 14 base pairs DNA duplex. B. Model for the binding of two $H\text{-}NS_{FL}$ C21S homodimers to one 20 base pairs DNA molecule. C. Model for the binding of two $H\text{-}NS_{FL}$ C21S homodimers to two 20 base pairs DNA molecules. D. Binding of a homotetramer of $H\text{-}NS_{FL}$ C21S binding to one 20 base pairs DNA molecule.

were measured using a DNA gel shift assay, as described in section 2.9.2. Figure 5-3 shows the DNA gel shift assay for the H-NS and StpA nucleic acid binding domains in the presence of 2 μM 20 base pairs DNA. Under the conditions tested (10 mM sodium phosphate pH 7.0, 100 mM NaCl, 10 μM NaN_3) no interaction is observed between the H-NS nucleic acid binding domain and the 20 base pairs DNA. However, when the StpA nucleic acid binding domain was used a band corresponding to a complex of $StpA_{91-134}$ bound to the DNA is observed, when the concentration of protein is 62.5 μM or greater. The proportion of the free DNA decreases as the $StpA_{91-134}$ concentration increases with the apparent K_d being approximately 1 mM.

5.4 ASSIGNMENT OF THE NUCLEIC ACID BINDING OF H-NS FROM *S. TYPHIMURIUM*

The gel shift assay results show that the *E. coli* StpA nucleic acid binding domain binds to the double stranded AAT20mer DNA molecule whereas the *S. typhimurium* H-NS_{90-137} protein does not bind. However, a DNA gel shift assay does not measure the interaction under equilibrium conditions. The DNA:protein complex is separated from the free DNA by electrophoresis and will dissociate, resulting in a lower proportion of protein in the band corresponding to the DNA:protein complex relative to the proportion of the total protein bound to the DNA in the sample prior to loading onto the gel. Therefore a DNA gel shift assay will make the interaction appear weaker than it actually is. To determine the binding affinity of the H-NS and StpA nucleic acid binding domains under equilibrium conditions, DNA titrations were performed using NMR spectroscopy. The DNA titrations allow a direct comparison of the binding affinity of the *S. typhimurium* H-NS_{90-137} and *E. coli* $StpA_{91-134}$ to be made and identify the amino acids that are involved in the DNA binding interface.

To be able to directly compare the binding of these two nucleic acid binding domains to the AAT20mer DNA duplex it is necessary to obtain the sequence specific backbone amide assignments for both *S. typhimurium* H-NS_{90-137} and *E. coli* $StpA_{91-134}$. The assignments and the solution structure of *E. coli* $StpA_{91-134}$ have been solved by Dr S. Ono and Dr M. Williams in this laboratory (unpublished results). The *E. coli* H-NS nucleic acid binding domain structure (residues Ala91 to Gln137) has been published (Shindo et al., 1995) as well as the assigned [^1H,^{15}N]-HSQC spectrum of *E. coli* H-NS_{61-137} (Shindo et al., 1999). However, these assignments could not be used to unambiguously assign all of the [^1H,^{15}N]-HSQC spectrum of *S. typhimurium* H-NS_{90-137} because there are five amino acid differences (A92Q, E124D, Q129S, E131D and E137Q) between the H-NS nucleic acid binding domains of *E. coli* and *S. typhimurium*

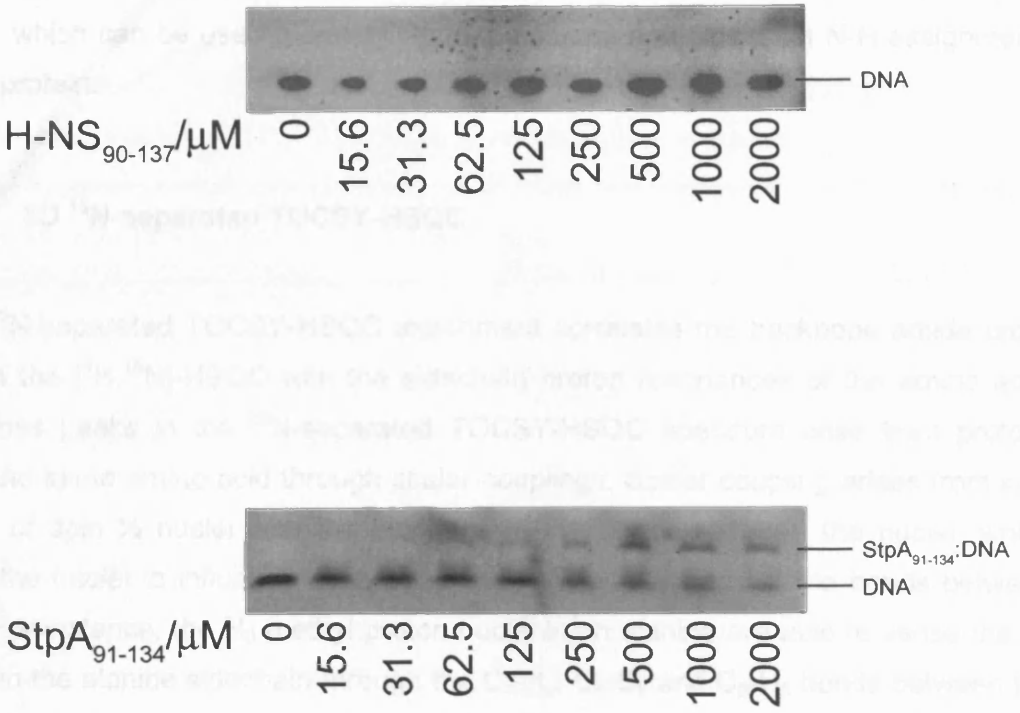


Figure 5-3. Gel shift assay for the binding of H-NS_{90-137} and $StpA_{91-134}$ to double stranded AAT20mer DNA (5'-GCAATAATAATAATACG-3'). Protein concentrations are indicated in μM below the appropriate lane of the gel.

and the buffer conditions used to determine the backbone amide assignments of H-NS_{61-137} and $StpA_{91-134}$ are different. Therefore, it was necessary to acquire additional NMR experiments; a 3D ^{15}N-separated TOCSY-HSQC and a ^{15}N-edited NOESY-HSQC, which can be used to establish the backbone and sidechain N-H assignments for the protein.

5.4.1 3D ^{15}N-separated TOCSY-HSQC

A 3D ^{15}N-separated TOCSY-HSQC experiment correlates the backbone amide cross peak in the [^{1}H,^{15}N]-HSQC with the sidechain proton resonances of the amino acid. The cross peaks in the ^{15}N-separated TOCSY-HSQC spectrum arise from protons within the same amino acid through scalar couplings. Scalar coupling arises from spin pairing of spin ½ nuclei with the electrons in the bonds between the nuclei, which allows the nuclei to influence each other through the electrons in the bonds between them. For instance, the H_β methyl proton nuclei in an alanine are able to sense the H_α proton in the alanine sidechain through the C_α-H_α, C_α-C_β and C_β-H_β bonds between the two protons. If the H_α nucleus is oriented with the bulk magnetic field of the NMR spectrometer the magnetic field experienced by H_β nuclei will be slightly stronger than the field experienced when H_α is aligned against the bulk magnetic field. The H_β methyl group nuclei magnetic resonance line is split by the H_α resulting in two resonance lines in the 1D 1H NMR spectrum

In a total correlation spectroscopy (TOCSY) experiment scalar coupling is used to transfer magnetisation between the proton nuclei to establish correlations between the protons within the same amino acid. The diagrammatic representation of a ^{1}H,^{1}H-TOCSY spectrum for an amino acid with a sidechain containing H_β and $H\gamma$ protons is shown in Figure 5-4. Cross peaks on the diagonal show the magnetic resonance frequency of the proton nuclei (NH, H_α, H_β or H_γ) in the amino acid. The cross peaks in the spectra either side of the diagonal line represent cross peaks produced due to magnetisation transfer by scalar coupling from one proton nuclei to the other. The pattern of cross peaks in the TOCSY spectrum can be used to identify the amino acid. Threonine, for example can be identified from the unique pattern of the H_α, H_β and H_γ cross peaks. The H_α and H_β cross peaks of threonine are unique in that both are found between 4 and 5 ppm. The H_β cross peak for the other amino acids is normally located between 1.5 and 4 ppm. However, amino acids such as glutamate, glutamine and methionine have equivalent spin systems and therefore give rise to the same pattern of cross peaks in the ^{1}H,^{1}H-TOCSY spectrum. To assign these amino acids correlations

need to be made with the neighbouring amino acids so that their position in the polypeptide chain can be determined. This will be described in more detail in section 5.4.3.

For a large macromolecule containing hundreds of protons it is necessary to separate the cross peaks in the $^{1}H,^{1}H$-TOCSY spectrum through scalar coupling to the ^{15}N atom in the backbone amide of the amino acid. The 3D ^{15}N-separated TOCSY-HSQC experiment generates a three dimensional spectrum of cross peaks, which contains only a subset of the cross peaks observed in a $^{1}H,^{1}H$-TOCSY spectrum (Figure 5-4). Each amino acid backbone amide is correlated to the other protons within the same amino acid but separated from the cross peaks of other amino acids in terms of the nitrogen chemical shift of the backbone amide.

5.4.2 ^{15}N-edited NOESY-HSQC

The ^{15}N-edited NOESY-HSQC NMR spectroscopy experiment detects the Nuclear Overhauser enhancement (NOE) connectivities between a backbone amide group and the spatially neighbouring protons, either within the same amino acid or in other residues. Cross peaks corresponding to dipole-dipole couplings (through space as opposed to being through chemical bonds as is the case for scalar coupling) between a backbone amide and protons within the same amino acid (i), or protons in the previous amino acid (i-1) in the sequence, are visible in the ^{15}N-edited NOESY-HSQC spectrum. The exclusive use of L-enantiomer amino acids and the tendency for a negative value of the phi backbone torsion angle, for α-helical and β-strand secondary structures (see Ramachandran plot in Figure 4-9), means that the i amide proton and the i+1 sidechain tend to point in opposite directions. This results in weak i amide proton to i+1 sidechain protons NOE's; corresponding cross peaks are usually not observed in the ^{15}N-edited ^{1}H-NOESY spectrum. Visible amide to i+1 sidechain NOE's are indicators of sharp left hand turns (positive phi angle) and are relatively rare, usually being associated with glycine and cis-proline residues.

5.4.3 Walking Along the Polypeptide Backbone

By overlaying the ^{15}N-edited NOESY-HSQC (Bax *et al.*, 1990), ^{15}N-separated TOCSY-HSQC (Marion *et al.*, 1989) and [$^{1}H,^{15}N$]-HSQC (Kay et al., 1992) spectra the NOE connectivities between neighbouring amino acids can be determined, allowing

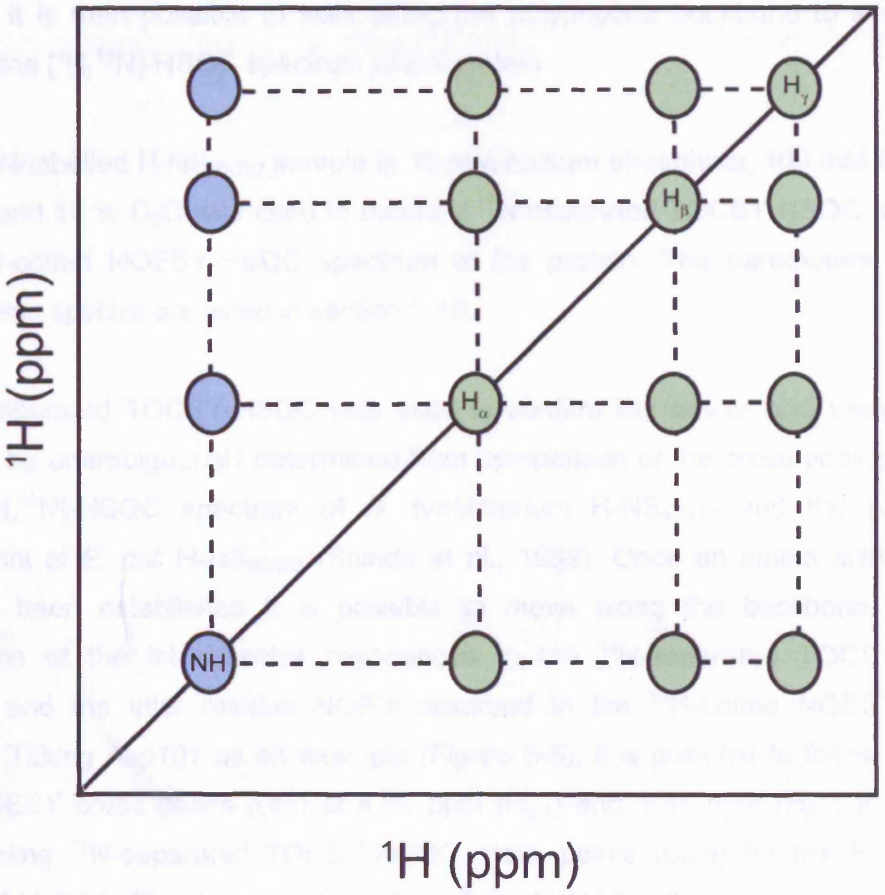


Figure 5-4. Representation of a ^{1}H,^{1}H-TOCSY spectrum for an amino acid showing the pattern of cross peaks observed for an amino acid with NH and H_α backbone protons and H_β and H_γ sidechain protons. All of the cross peaks shown in the diagram would be visible in a ^{1}H,^{1}H-TOCSY spectrum The pattern of cross peaks shown in blue indicate the cross peaks that are visible in a ^{15}N-separated TOCSY-HSQC spectrum.

assignment of the backbone amide and sidechain NH cross peaks in the [^{1}H,^{15}N]-HSQC spectrum of the protein. When the ^{15}N-separated TOCSY-HSQC and ^{15}N-edited NOESY-HSQC cross peaks overlay the cross peaks correspond to an intra residue resonance, whereas cross peaks that are only found in the ^{15}N-edited NOESY-HSQC spectra correspond to inter residue NOE's. The stronger, most intense of these inter residue NOE's arise from dipolar coupling between the protons in neighbouring amino acids. Having identified an amino acid type using the ^{15}N-separated TOCSY-HSQC spectrum, it is then possible to walk along the polypeptide backbone to assign the amides in the [^{1}H,^{15}N]-HSQC spectrum of the protein.

A 2 mM ^{15}N-labelled H-NS$_{90-137}$ sample in 10 mM sodium phosphate, 100 mM NaCl, 10 µM NaN_3 and 10 % D_2O was used to record a ^{15}N-separated TOCSY-HSQC spectrum and a ^{15}N-edited NOESY-HSQC spectrum of the protein. The parameters used to acquire these spectra are listed in section 2.10.

The ^{15}N-separated TOCSY-HSQC was used to confirm the amino acid assignments that could be unambiguously determined from comparison of the cross peak positions in the [^{1}H,^{15}N]-HSQC spectrum of *S. typhimurium* H-NS$_{90-137}$ and the published assignments of *E. coli* H-NS$_{60-137}$ (Shindo et al., 1999). Once an amino acid starting point has been established it is possible to move along the backbone using a combination of the intra-residue resonances in the ^{15}N-separated TOCSY-HSQC spectrum and the inter residue NOE's observed in the ^{15}N-edited NOESY-HSQC spectrum. Taking Asp101 as an example (Figure 5-5), it is possible to follow the ^{15}N-edited NOESY cross peaks (red) at 4.46 ppm ($H_{\alpha-1}$) and 1.91 ppm ($H_{\beta-1}$) to find the corresponding ^{15}N-separated TOCSY-HSQC cross peaks (blue) for the H_α and H_β protons of Val100. The H_α and H_β protons from Val100 will be correlated with the backbone amide proton of Val100. The nitrogen and proton chemical shifts of the Val100 amide proton, in the ^{15}N-TOCSY/NOESY spectra will be the same as the chemical shifts observed for the Val100 amide cross peak in the [^{1}H,^{15}N]-HSQC spectrum and so this amide cross peak can be assigned.

5.5 2D [^{1}H,^{15}N]-HSQC NMR SPECTRUM OF *S. TYPHIMURIUM* H-NS$_{90-137}$ AND *E. COLI* STPA$_{91-134}$

The [^{1}H,^{15}N]-HSQC spectra of 2 mM H-NS$_{90-137}$ and StpA$_{91-134}$ in 10 mM sodium phosphate pH 7.0, 100 mM NaCl, 10 µM NaN_3 was recorded using a Varian UNITYplus

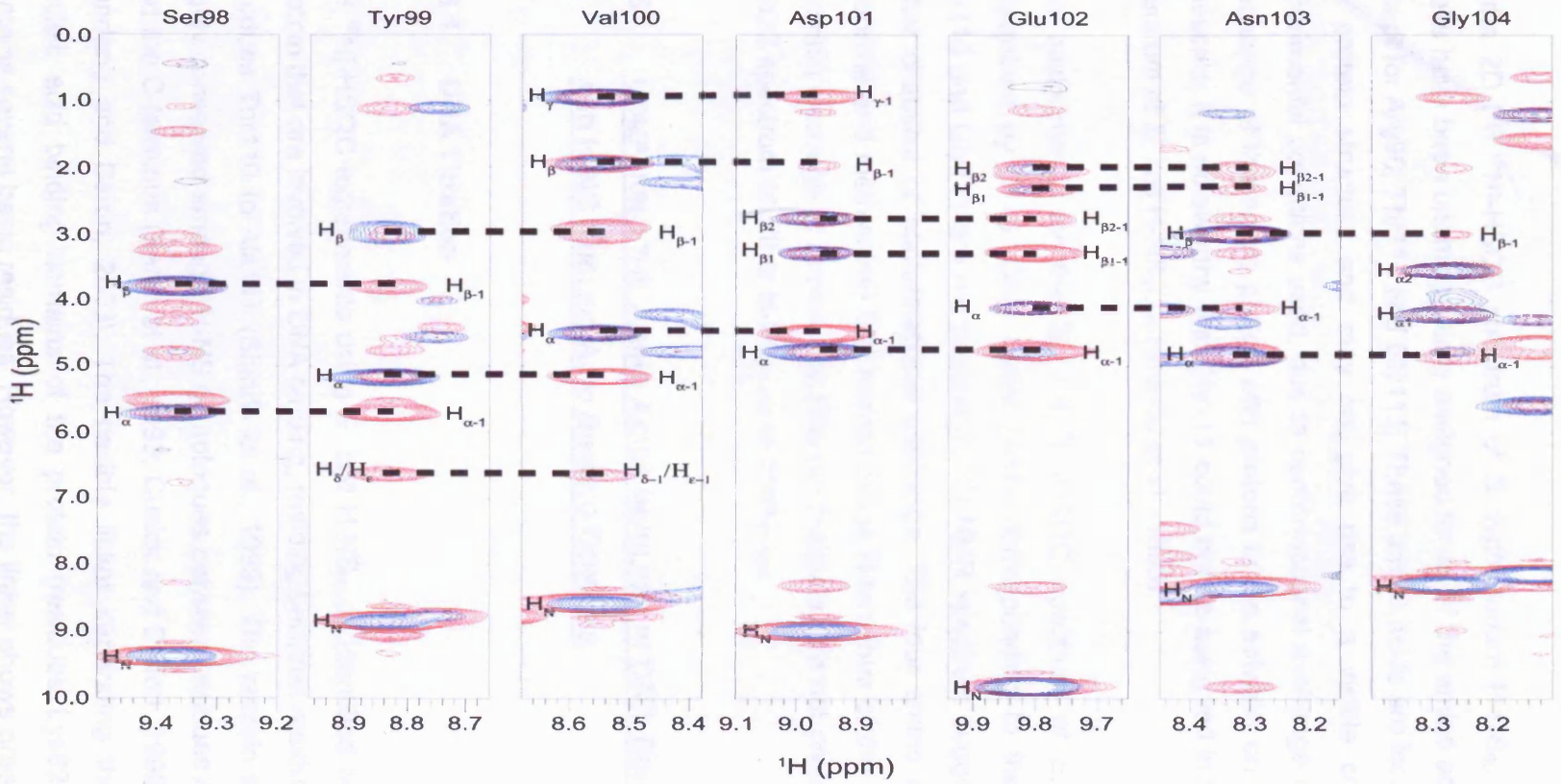


Figure 5-5. Overlaid ^{15}N-edited NOESY-HSQC (red) and ^{15}N-separated TOCSY-HSQC (blue) NMR spectra for H-NS$_{90\text{-}137}$. Each window shows the NOESY and TOCSY cross peaks corresponding to the ^{15}N chemical shift of the backbone amide proton of the amino acid indicated. The dashed lines indicate alignment between the chemical shift of protons (H_α, H_β, H_γ, H_δ or H_ε) of the amino acid selected and an NOE cross peak in the NOESY spectra corresponding to the i+1 amino acid in the polypeptide chain.

500 MHz spectrometer. As described in section 3.6, the 2D [^{1}H,^{15}N]-HSQC NMR spectrum shows the backbone amide resonances for each amino acid (except proline) in the protein and some of the sidechain N-H groups.

In the 2D [^{1}H,^{15}N]-HSQC spectrum of *S. typhimurium* H-NS$_{90\text{-}137}$ (Figure 5-6) cross peaks have been unambiguously assigned for all of the amino acids in the polypeptide except for Arg90, Thr110 and Gly111. These amino acids are located in loop regions of the protein structure and may not give rise to a visible cross peak under the experimental conditions used, due to conformational exchange or chemical exchange (exchange of the amide proton with protons in the solvent) on an intermediate NMR timescale. It is noteworthy that Gly111 could not be assigned in the 2D [^{1}H,^{15}N]-HSQC spectrum of *E. coli* H-NS$_{91\text{-}137}$ (Shindo et al., 1995).

The assignments for the 2D [^{1}H,^{15}N]-HSQC spectrum of *E. coli* StpA$_{91\text{-}134}$ were determined by Dr S. Ono. Cross peaks corresponding to the backbone amide of Thr110 and Gly111 are not present in this NMR spectrum suggesting they are also in either chemical or conformational exchange. The four amino acids (glycine, serine, threonine and methionine) that remain at the N-terminus of the protein construct after thrombin cleavage to remove the His$_6$-tag residues, are not present in the 2D [^{1}H,^{15}N]-HSQC spectrum of either H-NS$_{90\text{-}137}$ or StpA$_{91\text{-}134}$.

5.6 Comparing the Amino Acids Involved in DNA Binding Between StpA and H-NS Nucleic Acid Binding Domains

5.6.1 DNA Titration

[^{1}H,^{15}N]-HSQC experiments using *E. coli* H-NS$_{61\text{-}137}$ identified two loop regions in the protein that are involved in DNA binding, residing between residues Ala80 to Lys96 and residues Thr110 to Ala117 (Shindo et al., 1999). The nucleic acid binding domain is highly conserved amongst H-NS homologues between residues Arg90 (in *E. coli* H-NS) and the C-terminus (Bertin et al., 1999; Cusick and Belfort, 1998; Dorman et al., 1999; Tendeng and Bertin, 2003). The flexible linker separating the oligomerisation and nucleic acid binding domains of the protein (residues Lys82 to Lys89 in *E. coli*) contains several basic residues. However, the linker shows greater sequence variation than the residues within the nucleic acid binding domain.

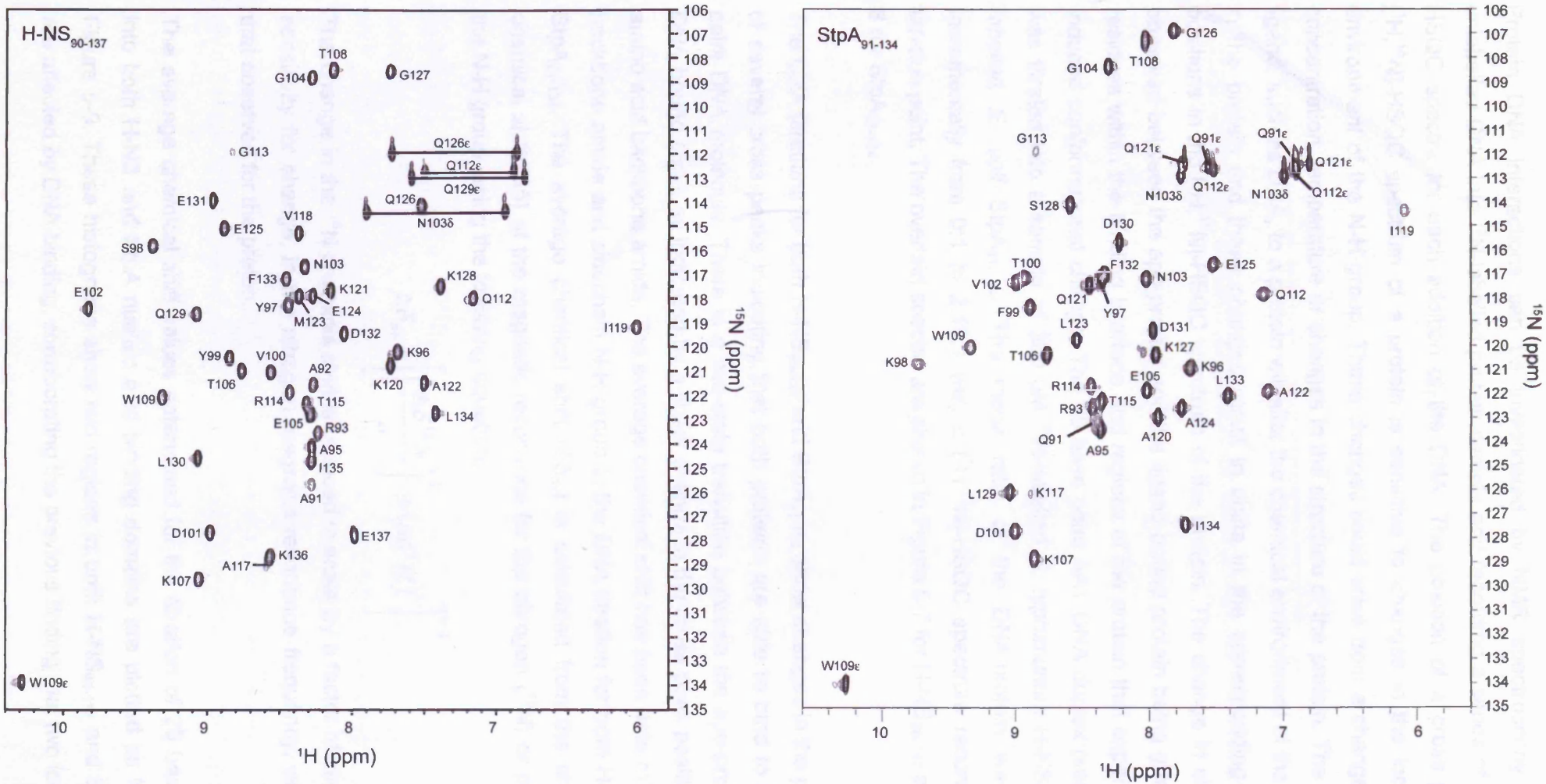


Figure 5-6. Annotated 2D [^{1}H,^{15}N]-HSQC spectra of *S. typhimurium* H-NS$_{90-137}$ (left) *E. coli* StpA$_{91-134}$ (right) and recorded using a Varian UNITYplus 500 MHz NMR spectrometer at 25°C. The sample contained 2 mM ^{15}N-labelled *S. typhimurium* H-NS$_{90-137}$ dissolved in 10 mM sodium phosphate pH 7.0, 100 mM NaCl, 10 µM NaN_3, 10% D_2O. Cross peaks are labelled with the corresponding amino acid identity.

Protein DNA interactions can be investigated by NMR spectroscopy by titrating unlabelled DNA into ^{15}N-labelled protein sample and recording a series of 2D [^{1}H,^{15}N]-HSQC spectra for each addition of the DNA. The position of a cross peak in the [^{1}H,^{15}N]-HSQC spectrum of a protein is sensitive to changes in the local chemical environment of the N-H group. These changes could arise from a change in pH, salt concentration, temperature or changes in the structure of the protein. The binding of a ligand, such as DNA, to a protein will alter the chemical environment of the N-H groups in the protein and these changes result in shifts in the corresponding cross peak positions in the [^{1}H,^{15}N]-HSQC spectrum of the protein. The change in chemical shift observed between the apo-protein and the ligand bound protein being greatest for the residues within the binding interface and regions of the protein that experience ligand induced conformational change. The 20 base pairs AAT DNA duplex (see Table 5-1) was titrated into a sample of 300 μM ^{15}N-labelled *S. typhimurium* H-NS$_{90\text{-}137}$ or ^{15}N-labelled *E. coli* StpA$_{91\text{-}134}$. The molar ratio of the DNA:protein was increased incrementally from 0:1 to 2.19:1 with a [^{1}H,^{15}N]-HSQC spectrum recorded at each titration point. The overlaid spectra are shown in Figure 5-7 for H-NS$_{90\text{-}137}$ and Figure 5-8 for StpA$_{91\text{-}134}$.

The DNA titrations for both H-NS$_{90\text{-}137}$ and StpA$_{91\text{-}134}$ show changes in the peak position of several cross peaks indicating that both proteins are able to bind to the 20 base pairs DNA molecule. There is a two-state transition between the apo-protein and the DNA bound state, as indicated by a linear change in the cross peak positions for each amino acid backbone amide. The average chemical shift has been determined for each backbone amide and sidechain N-H group in the DNA titration for both H-NS$_{90\text{-}137}$ and StpA$_{91\text{-}134}$. The average chemical shift ($\Delta\delta_{av}$) is calculated from the change in the chemical shift ($\Delta\delta$) of the magnetic resonance for the nitrogen (^{15}N) or proton (^{1}H) of the N-H group, using the following equation.

$$\Delta\delta_{av} = \left[\left(\frac{\Delta\delta^{15}N}{6}\right)^2 + \left(\Delta\delta^{1}H\right)^2\right]^{1/2}$$

The change in the ^{15}N chemical shift is reduced in scale by a factor of six to make the sensitivity for change, in the nitrogen magnetic resonance frequency, comparable to that observed for the proton.

The average chemical shift values determined for the titration of 20 base pairs DNA into both H-NS and StpA nucleic acid binding domains are plotted as histograms in Figure 5-9. These histograms show two regions in both H-NS$_{90\text{-}137}$ and StpA$_{91\text{-}134}$ that are affected by DNA binding, corroborating the previous finding that two loop regions in

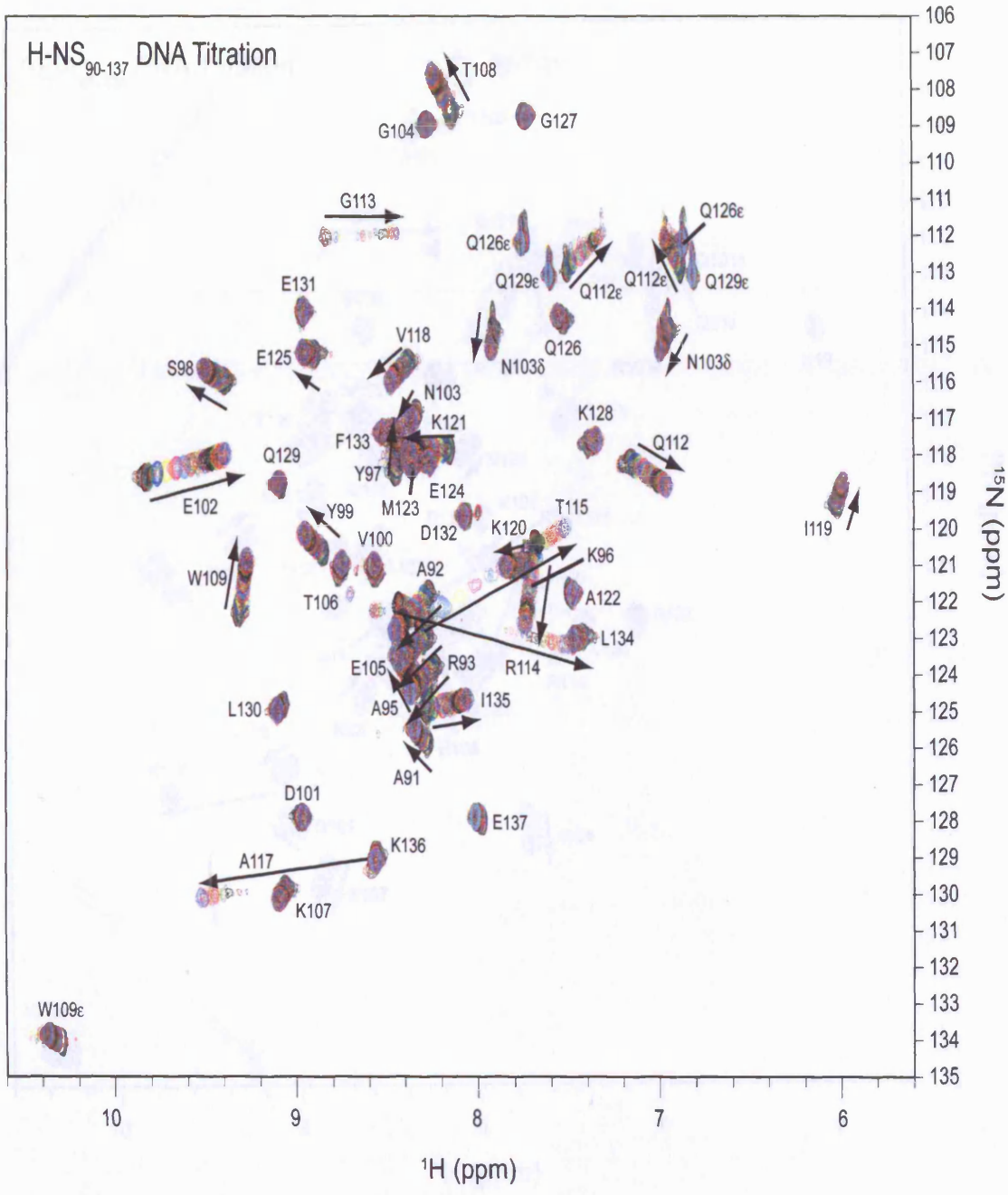


Figure 5-7. DNA titration for H-NS_{90-137}. Overlaid 2D [^{1}H,^{15}N]-HSQC spectra of samples containing [^{1}H,^{15}N] labelled *E. coli* H-NS_{90-137} in 10 mM sodium phosphate pH 7.0, 100 mM NaCl, 10 µM sodium azide, 10% D_2O. The following molar ratios of DNA:H-NS_{90-137} are shown: 0:1 (Black), 0.005:1 (Red), 0.01:1 (Green), 0.04:1 (Blue), 0.08:1 (Yellow), 0.12:1 (Magenta), 0.16:1 (Cyan), 0.20:1 (Brown), 0.24:1 (Chartreuse), 0.28:1 (Coral), 0.32:1 (Dark Blue), 0.37:1 (DarkRed), 0.41:1 (DarkGreen), 0.45:1 (DodgerBlue), 0.51:1 (ForestGreen) 0.57:1 (Gold), 0.67:1 (Red), 0.77, (Indian Red), 0.97:1 (Orange), 1.58:1 (Purple), 2.19:1 (Turquoise).

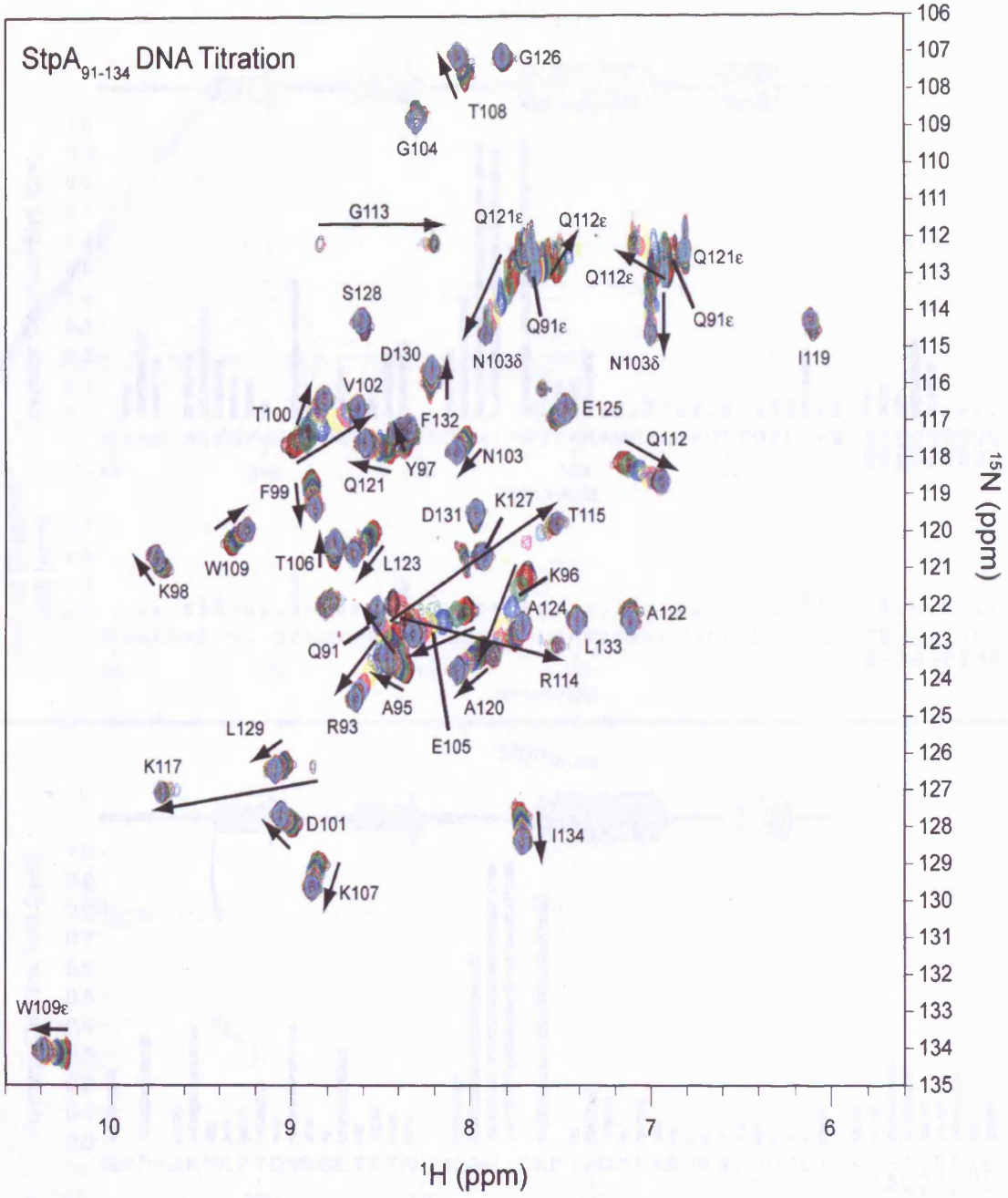


Figure 5-8. DNA titration for $StpA_{91-134}$. Overlaid 2D [^{1}H,^{15}N]-HSQC spectra of samples containing ^{1}H,^{15}N labelled *S. typhimurium* $StpA_{91-134}$ in 10 mM sodium phosphate pH 7.0, 100 mM NaCl, 10 µM sodium azide, 10% D_2O. The following molar ratios of DNA:$StpA_{91-134}$ are shown: 0:1 (Black), 0.01:1 (Red), 0.05:1 (Green), 0.13:1 (Blue), 0.20:1 (Yellow), 0.28:1 (Magenta), 0.35:1 (Cyan), 0.43:1 (Brown), 0.50:1 (Chartreuse), 0.58:1 (Coral), 0.66:1 (Dark Blue), 0.73:1 (DarkRed), 0.81:1 (DarkGreen), 0.88:1 (DodgerBlue), 0.96:1 (ForestGreen) 1.03:1 (Gold), 1.29:1 (IndianRed), 1.74:1 (Purple), 2.80:1 (Turquoise).

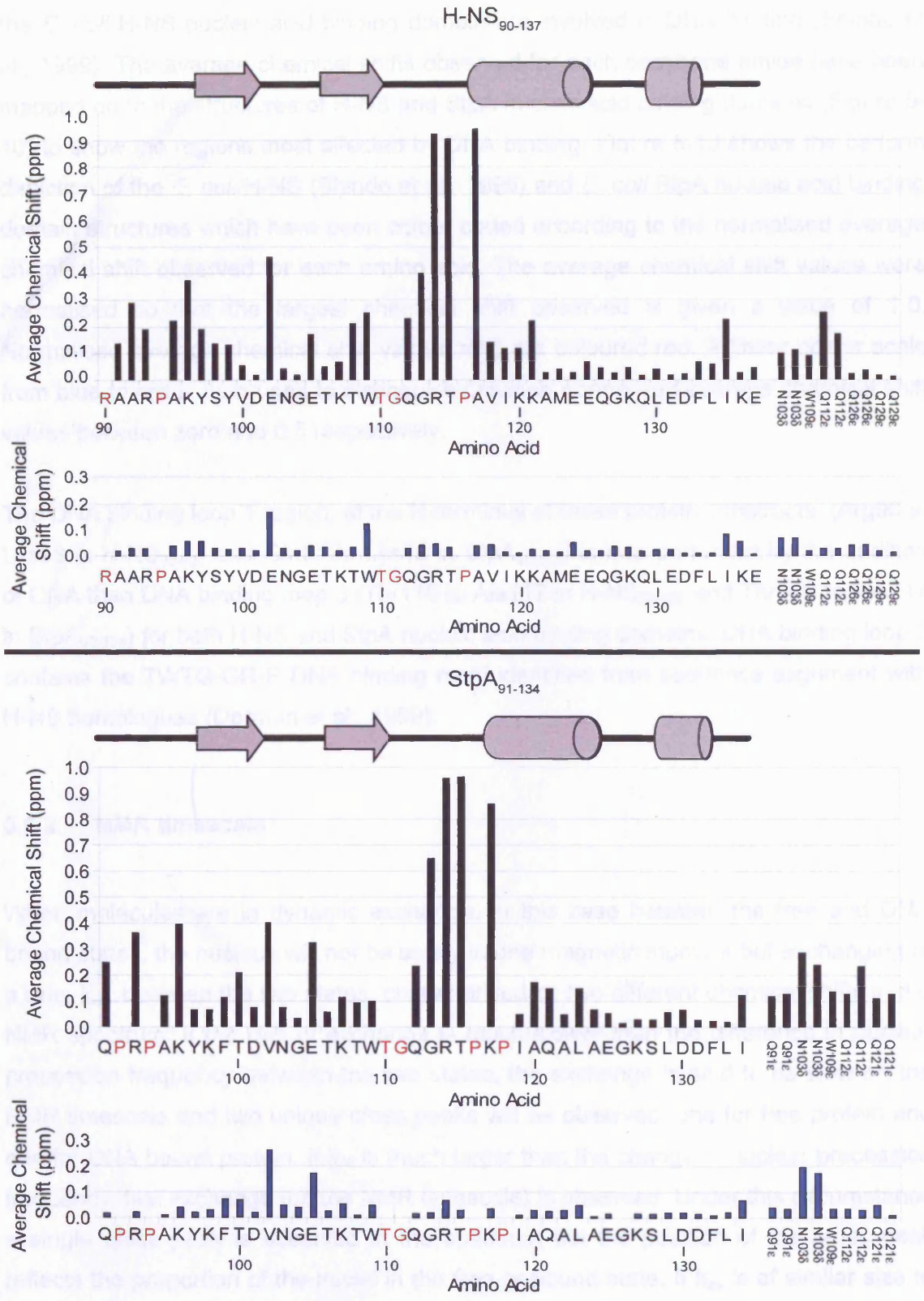


Figure 5-9. Histograms of the average chemical shift changes induced upon formation of a complex with a 20 base pairs DNA duplex (black). The average chemical shift observed when 438 mM NaCl (H-NS_{90-137}) or 475 mM NaCl ($StpA_{91-134}$) is added to the apo protein sample is shown (blue). The results for H-NS_{90-137} are shown above the results for $StpA_{91-134}$. The amino acids coloured red are not present in the 2D [^{1}H,^{15}N]-HSQC spectrum so the average chemical shift changes for these residues could not be determined.

the *E. coli* H-NS nucleic acid binding domain are involved in DNA binding (Shindo et al., 1999). The average chemical shifts observed for each backbone amide have been mapped on to the structures of H-NS and StpA nucleic acid binding domains (Figure 5-10) to show the regions most affected by DNA binding. Figure 5-10 shows the cartoon depiction of the *E. coli* H-NS (Shindo et al., 1995) and *E. coli* StpA nucleic acid binding domain structures which have been colour coded according to the normalised average chemical shift observed for each amino acid. The average chemical shift values were normalised so that the largest chemical shift observed is given a value of 1.0. Normalised average chemical shift values >0.5 are coloured red. A linear colour scale from blue to red is then used to distinguish between normalised average chemical shift values between zero and 0.5 respectively.

The DNA binding loop 1 region, at the N-terminus of these protein constructs, (Arg90 to Lys98 in H-NS_{90-137} and Gln91 to Lys98 in $StpA_{91-134}$) is less perturbed by the addition of DNA than DNA binding loop 3 (Thr110 to Ala117 in H-NS_{90-137} and Thr110 to Lys117 in $StpA_{90-134}$) for both H-NS and StpA nucleic acid binding domains. DNA binding loop 3 contains the TWTG-GR-P DNA binding motif identified from sequence alignment with H-NS homologues (Dorman et al., 1999).

5.6.2 NMR timescale

When molecules are in dynamic exchange, in this case between the free and DNA bound states, the nucleus will not be solely in one magnetic moment but exchanging at a rate, k_{ex}, between the two states, characterised by two different chemical shifts in the NMR spectrum. If the rate of exchange is much slower than the difference in nuclear precession frequency between the two states, the exchange is said to be slow on the NMR timescale and two unique cross peaks will be observed, one for free protein and one for DNA bound protein. If k_{ex} is much larger than the change in nuclear precession frequency, fast exchange (on the NMR timescale) is observed. Under this circumstance a single cross peak is observed in the spectrum but the position of the cross peak reflects the proportion of the nuclei in the free or bound state. If k_{ex} is of similar size to the change in nuclear precession frequency then the process is said to be in intermediate exchange on the NMR timescale and the peaks representing the mid point of the titration (where an average of 50% of the protein is in the bound state) will be broadened and have reduced intensity. Figure 5-11A shows the different cross peak

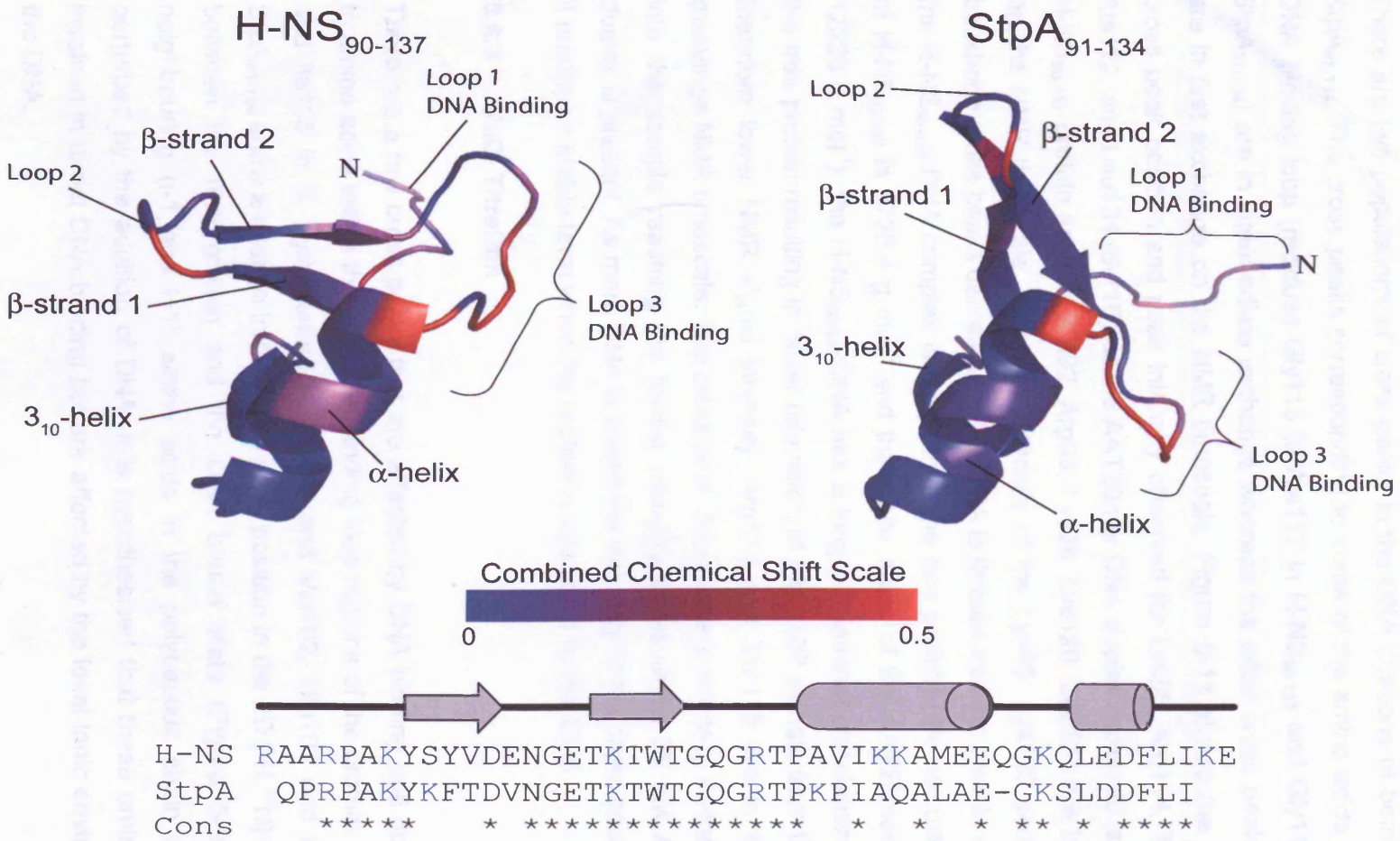


Figure 5-10. Cartoon representations of the H-NS$_{90-137}$ and StpA$_{91-134}$ nucleic acid binding domains. The amino acids have been colour coded according to the average chemical shift observed between the free protein and the protein bound to a 20 base pairs DNA duplex. The amino acid sequences of H-NS$_{90-137}$ and StpA$_{91-134}$ are aligned with the secondary structure elements and positively charged amino acids are coloured blue.

profiles observed when the exchange process is in fast, intermediate or slow exchange on the NMR timescale.

There are two populations of cross peaks in the DNA titrations of both H-NS_{90-137} and $StpA_{91-134}$. The cross peaks corresponding to some of the amino acids within the main DNA binding loop (residues Gly113 to Ala117 in H-NS_{90-137} and Gly113 to Lys117 in $StpA_{91-134}$) are in intermediate exchange whereas the other cross peaks in the protein are in fast exchange on the NMR timescale. Figure 5-11 shows the changes in the cross peak position and peak intensity observed for Lys96, Arg114, Thr115, Lys120, Ala122, and Leu134 (5-11B) as the AAT20mer DNA duplex molecule is titrated into the H-NS_{90-137} protein sample. Ala92, Arg93, Lys96, Lys120, Leu134 are in fast exchange on the NMR timescale. The peak intensity of the Lys96, Lys120 and Leu134 amide backbone cross peaks decrease as the DNA is titrated into the protein sample because the H-NS_{90-137}:DNA complex is larger than the free protein (the calculated molar mass of H-NS_{90-137} is 5726.4 g mol^{-1} and the molar mass of the AAT20mer DNA duplex is 12228 g mol^{-1}); the H-NS_{90-137}:DNA has a longer rotational correlation time relative to the free protein resulting in faster relaxation of the NMR signals from the complex and therefore lower NMR signal intensity. Arg114 and Thr115 are in the intermediate exchange NMR timescale. The cross peak decreases in intensity as the DNA is titrated into the sample reaching the lowest intensity value when 36 µM AAT20mer DNA duplex is present. As more DNA is added the intensity of the cross peak increases until it reaches a stable level when the protein is saturated by the DNA.

5.6.3 NaCl Titration

There are a few cross peaks that are affected by DNA binding that do not correspond to amino acids within the two DNA binding loop regions of the protein. Glu102, Glu105 and Ile135 in *S. typhimurium* H-NS_{90-137} and Val102, Glu105 and Ile134 in *E. coli* $StpA_{91-134}$ show a change in the cross peak position in the 2D [^{1}H,^{15}N]-HSQC spectrum between the free protein and the DNA bound state (Figure 5-9). However the neighbouring (i-1 and i+1) amino acids in the polypeptide chain are only slightly perturbed by the addition of DNA. It is hypothesised that these amino acids are not involved in direct DNA binding but are affected by the local ionic environment close to the DNA.

To investigate the effect of an increased ionic environment on the H-NS and StpA nucleic acid binding domains separate NMR titrations were performed, where the sodium chloride content of the solvent was increased. At each NaCl concentration in

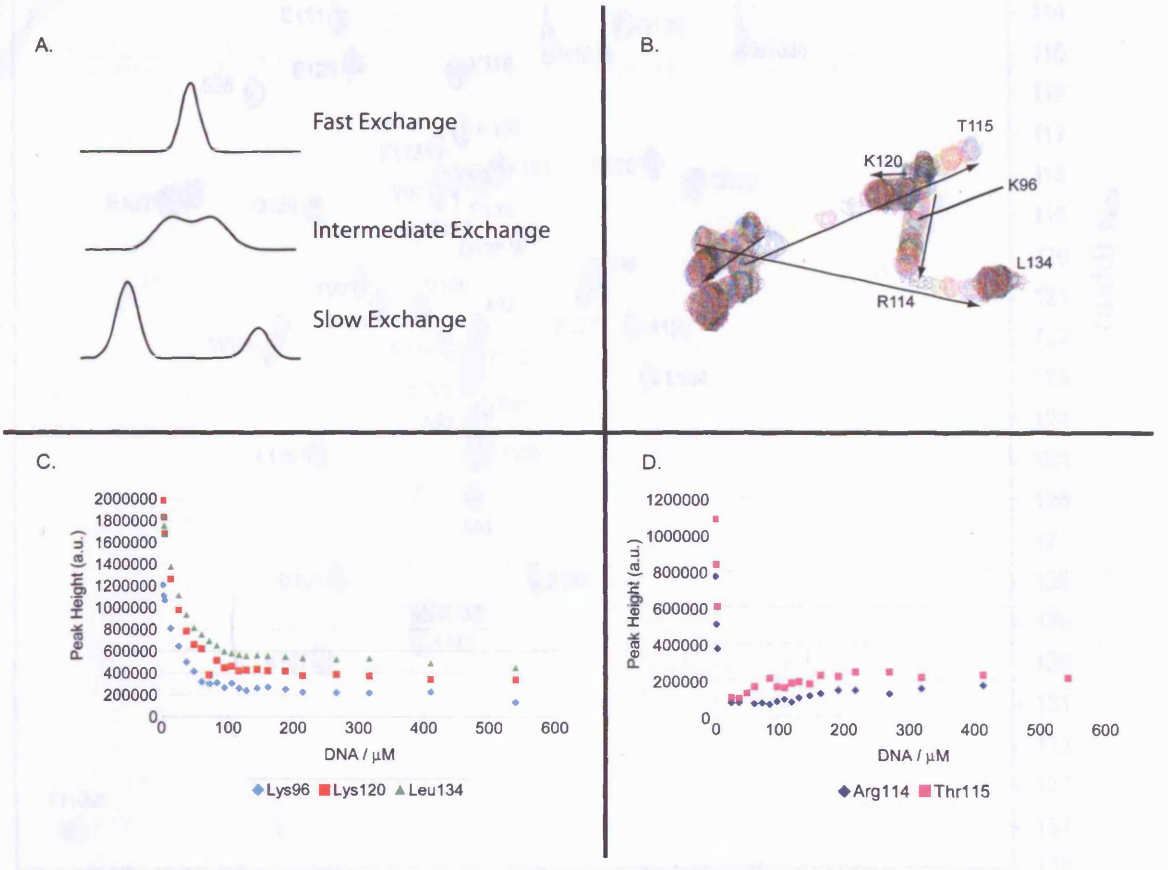


Figure 5-11. A. The effect of fast, intermediate or slow NMR timescale exchange on the resonance peaks observed in the NMR spectrum. B. View of the cross peak shifts for the amide backbone resonances of Lys96, Arg114, Arg115, Lys120 and Leu134. C. The change in peak intensity of the Lys96, Lys120 and Leu134 backbone amide cross peaks of H-NS_{90-137} as AAT20mer duplex DNA is titrated into the protein sample. D. The change in peak intensity of the Arg114 and Thr115 backbone amide cross peaks of H-NS_{90-137} as AAT20mer duplex DNA is titrated into the protein sample.

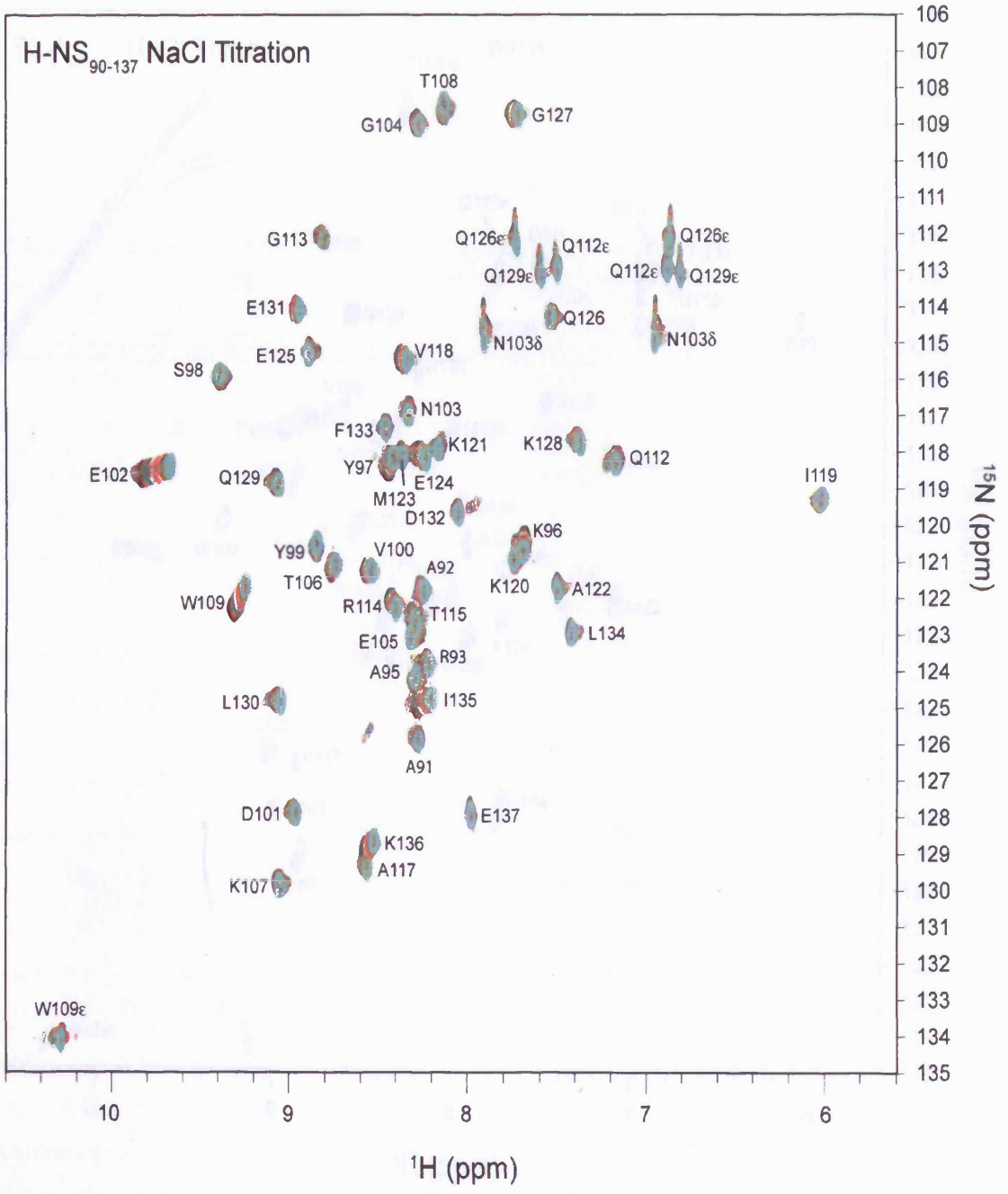


Figure 5-12. NaCl titration for H-NS_{90-137}. Overlaid 2D [^{1}H,^{15}N]-HSQC spectra of samples containing [^{1}H,^{15}N]-labelled *S. typhimurium* H-NS_{90-137} in 10 mM sodium phosphate pH 7.0, 10 μM sodium azide, 10% D_2O and various NaCl concentrations. The following NaCl concentrations are shown: 100 mM (Black), 139 mM (Red), 177 mM (Green), 208 mM (Blue), 248 mM (Yellow), 314 mM (Magenta), 376 mM (Cyan), 433 mM (Brown), 468 mM (Chartreuse), 538 mM (ForestGreen).

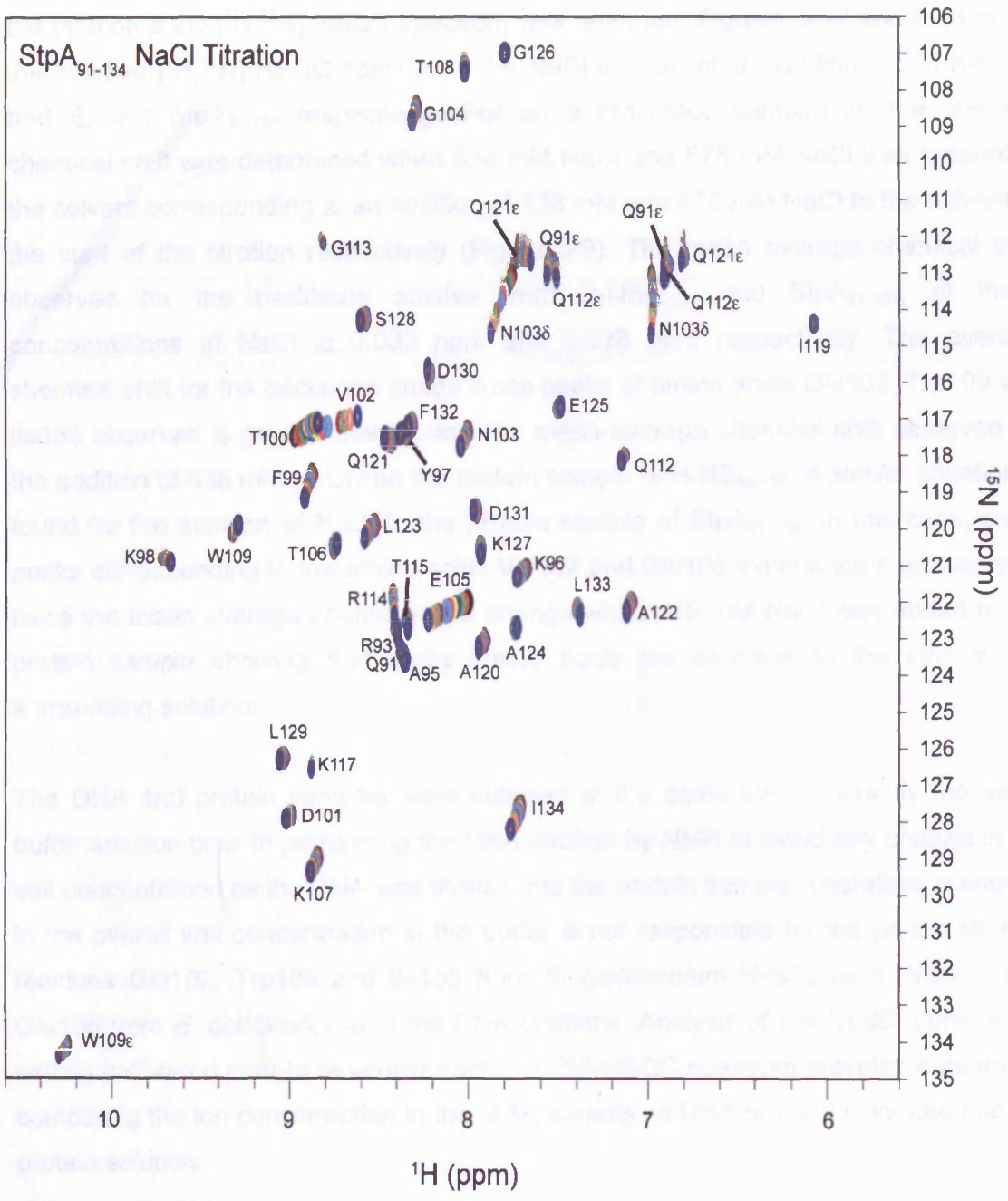


Figure 5-13. NaCl titration for $StpA_{91-134}$. Overlaid 2D [^{1}H,^{15}N]-HSQC spectra of samples containing [^{1}H,^{15}N]-labelled *E. coli* $StpA_{91-134}$ in 10 mM sodium phosphate pH 7.0, 10 µM sodium azide, 10% D_2O and various NaCl concentrations. The following NaCl concentrations are shown: 100 mM (Black), 145 mM (Red), 186 mM (Green), 224 mM (Blue), 258 mM (Yellow), 318 mM (Magenta), 383 mM (Cyan), 575 mM (Brown), 670 mM (Chartreuse), 813 mM (Coral), 1050 mM (Dark Blue).

the titration a 2D [^{1}H,^{15}N]-HSQC spectrum was recorded. Figures 5-12 and 5-13 show the overlaid [^{1}H,^{15}N]-HSQC spectra for the NaCl titration of *S. typhimurium* H-NS$_{90-137}$ and *E. coli* StpA$_{91-134}$ respectively. For an approximate comparison the average chemical shift was determined when 538 mM NaCl and 575 mM NaCl was present in the solvent corresponding to an addition of 438 mM and 475 mM NaCl to the solvent at the start of the titration respectively (Figure 5-9). The mean average chemical shift observed for the backbone amides from H-NS$_{90-137}$ and StpA$_{91-134}$, at these concentrations of NaCl is 0.033 ppm and 0.038 ppm respectively. The average chemical shift for the backbone amide cross peaks of amino acids Glu102, Trp109 and Ile135 observed is greater than double the mean average chemical shift observed for the addition of 438 mM NaCl into the protein sample of H-NS$_{90-137}$. A similar situation is found for the addition of NaCl to the protein sample of StpA$_{91-134}$. In this case, cross peaks corresponding to the amino acids Val102 and Glu105 experience a greater than twice the mean average chemical shift change when 475 mM NaCl was added to the protein sample showing that these amino acids are sensitive to the ions in the surrounding solution.

The DNA and protein samples were dialysed at the same time in exactly the same buffer solution prior to performing the DNA titration by NMR to avoid any change in the salt concentration as the DNA was titrated into the protein sample. Therefore, a change in the overall salt concentration in the buffer is not responsible for the perturbation of residues Glu102, Trp109 and Ile135 from *S. typhimurium* H-NS$_{90-137}$ or Val102 and Glu105 from *E. coli* StpA$_{91-134}$ in the DNA titrations. Analysis of the ^{1}H 90° pulse width setting calibrated prior to recording each [^{1}H,^{15}N]-HSQC spectrum provides a means of comparing the ion concentration in the NMR sample as DNA or NaCl is titrated into the protein solution.

The ^{1}H 90° pulse width is the length of the radiowave pulse required to rotate the proton nuclei magnetic moment through 90°. If the concentration of ions in the NMR sample increases then more of the radiowaves will be absorbed by the ions so that a longer pulse width will be required to rotate the protons in the sample through 90°. Before each [^{1}H,^{15}N]-HSQC spectrum is recorded the ^{1}H 90° pulse width (pwh) is calibrated. The effect of adding DNA to the NMR samples of H-NS$_{90-137}$ and StpA$_{91-134}$ can be compared to the effect of adding NaCl by analysing the ^{1}H 90° pulse width used for each [^{1}H,^{15}N]-HSQC experiment.

The ^{1}H 90° pulse width setting used when the [^{1}H,^{15}N]-HSQC was acquired for both H-NS$_{90-137}$ and StpA$_{91-134}$ prior to the addition of DNA was 6.875 showing that the salt

concentration in these two NMR samples was the same. When the DNA was added to the ^{15}N-labelled H-NS$_{90-137}$ protein sample the ^{1}H 90° pulse width had to be increased throughout the titration (Figure 5-14). When the final DNA concentration of 540 µM AAT20mer DNA duplex was present the ^{1}H 90° pulse width was set to 7.175 µs. For comparison the ^{1}H 90° pulse width when 527 µM AAT20mer DNA duplex was added to StpA$_{91-134}$ was 7.175 µs showing that the ion concentration in the NMR sample was increased by the same amount in the H-NS$_{90-137}$ and StpA$_{91-134}$ DNA titrations for a comparable addition of DNA.

Figure 5-14 compares the increase in the ^{1}H 90° pulse width as the DNA was added to 300 µM *S. typhimurium* H-NS$_{90-137}$ with the ^{1}H 90° pulse width setting required when NaCl was added to the H-NS$_{90-137}$ protein sample. The ^{1}H 90° pulse width setting for the acquisition of the [^{1}H,^{15}N]-HSQC spectra of ^{15}N-labelled H-NS$_{90-137}$ increased from 6.875 µs to 7.175 µs as the DNA was added to the NMR sample. The change in the ^{1}H 90° pulse width required as the NaCl concentration in the sample is increased from 100 mM to 538 mM is from 6.925 µs to 8.725 µs (Figure 5-14). The ^{1}H 90° pulse width settings for the titration of DNA and NaCl into the ^{15}N-labelled StpA$_{91-134}$ protein samples are not shown because these titrations were performed using different NMR spectrometers so a direct comparison of the ^{1}H 90° pulse width settings between the DNA and NaCl titrations cannot be made.

The perturbation of the cross peak positions of Glu102, Trp109 and Ile135 from *S. typhimurium* H-NS$_{90-137}$ and Val102 and Glu105 from *E. coli* StpA$_{91-134}$ when the AAT20mer DNA duplex was added to the protein solution cannot be explained by an overall increase in the concentration of NaCl in the buffer solution. The ^{1}H 90° pulse width increased to 7.175 µs when 540 µM or 527 µM AAT20mer DNA duplex was added to H-NS$_{90-137}$ and StpA$_{91-134}$ respectively but this change in the ^{1}H 90° pulse width is equivalent to the addition of only 40 mM NaCl to the protein sample. However, as the perturbation of these amino acids can be at least partially replicated by the addition of approximately 400 mM NaCl this suggests that a high local concentration of ions close to the surface of the DNA could be responsible for the change in chemical shift of these backbone amide cross peaks (Glu102, Trp109 and Ile135 in H-NS$_{90-137}$ and Val102 and Glu105 in StpA$_{91-134}$) in the DNA titration and not the direct interaction of these amino acids with the DNA itself. If there was a direct interaction of these amino acids with the DNA you would expect to see an effect on the neighbouring amino acids as well. However, the addition of NaCl into the protein samples of H-NS$_{90-137}$ and StpA$_{91-134}$ demonstrates that a change in the ionic environment can also be responsible for the isolated perturbation of these amino acids.

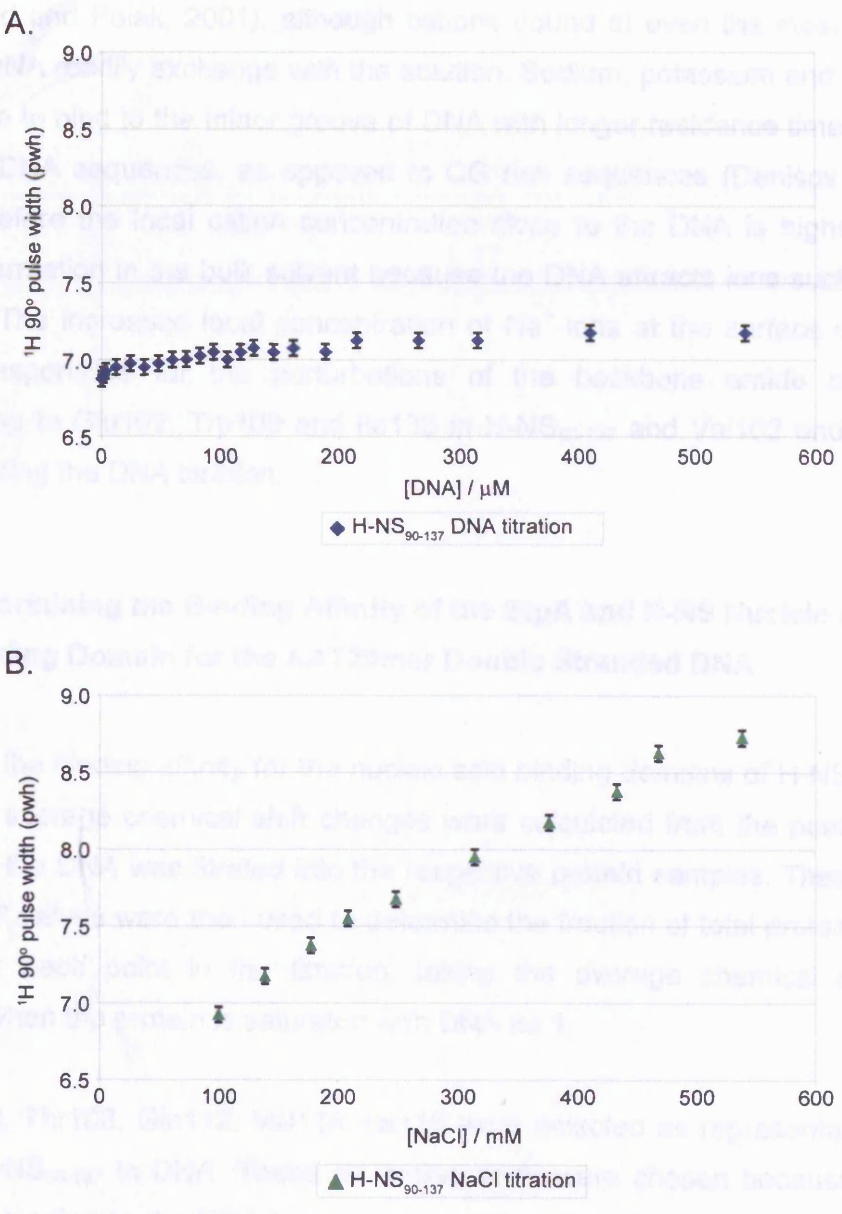


Figure 5-14. A. The ^{1}H 90° pulse width setting used for each DNA addition to the ^{15}N-labelled H-NS$_{90-137}$ protein sample containing 10 mM sodium phosphate, 100 mM NaCl, 10 µM NaN_3 and 10 % D_2O. B. The ^{1}H 90° pulse width setting used for the acquisition of the [^{1}H,^{15}N]-HSQC spectra of H-NS$_{90-137}$ when the NaCl concentration in the NMR sample was increased from 100 mM to 538 mM.

DNA has a highly charged phosphate backbone that is neutralised in solution by mobile counterions. Monovalent and divalent cations bind to both the major and minor groove of DNA (Hud and Polak, 2001), although cations bound at even the most favourable sites on B-DNA readily exchange with the solution. Sodium, potassium and ammonium ions are able to bind to the minor groove of DNA with longer residence times observed for AT rich DNA sequences, as opposed to CG rich sequences (Denisov and Halle, 2000). Therefore the local cation concentration close to the DNA is higher than the cation concentration in the bulk solvent because the DNA attracts ions such as Na^+ to its surface. The increased local concentration of Na^+ ions at the surface of the DNA could be responsible for the purturbations of the backbone amide cross-peaks corresponding to Glu102, Trp109 and Ile135 in H-NS_{90-137} and Val102 and Glu105 in $StpA_{91-134}$ during the DNA titration.

5.6.4 Determining the Binding Affinity of the StpA and H-NS Nucleic Acid Binding Domain for the AAT20mer Double Stranded DNA

To establish the binding affinity for the nucleic acid binding domains of H-NS and StpA for DNA the average chemical shift changes were calculated from the peak positions observed as the DNA was titrated into the respective protein samples. These average chemical shift values were then used to determine the fraction of total protein bound to DNA (Y) for each point in the titration, taking the average chemical shift value determined when the protein is saturated with DNA as 1.

Ser98, Tyr99, Thr108, Gln112, Val118, Ile119 were selected as representative of the binding of H-NS_{90-137} to DNA. These six amino acids were chosen because they are shifted upon binding to the DNA but are not greatly affected in the NaCl titration and they do not overlap with other cross-peaks so that the position of the cross peak at each DNA point in the DNA titration is well defined. Another reason for selecting these amino acids is because the cross peaks are in fast exchange on the NMR timescale so the cross peak position gives a reliable measure of the fraction of protein bound to DNA in the sample for each point in the titration. The residues in DNA binding loop 3, which are most affected by DNA binding, are in exchange on an intermediate NMR timescale; their cross peak positions are less well defined for the mid-points of the titration making the fitting of the peak position with respect to DNA concentration less reliable for determining the dissociation constant of the interaction.

Figure 5-15 shows the DNA binding isotherms for H-NS_{90-137} (residues Ser98, Tyr99,

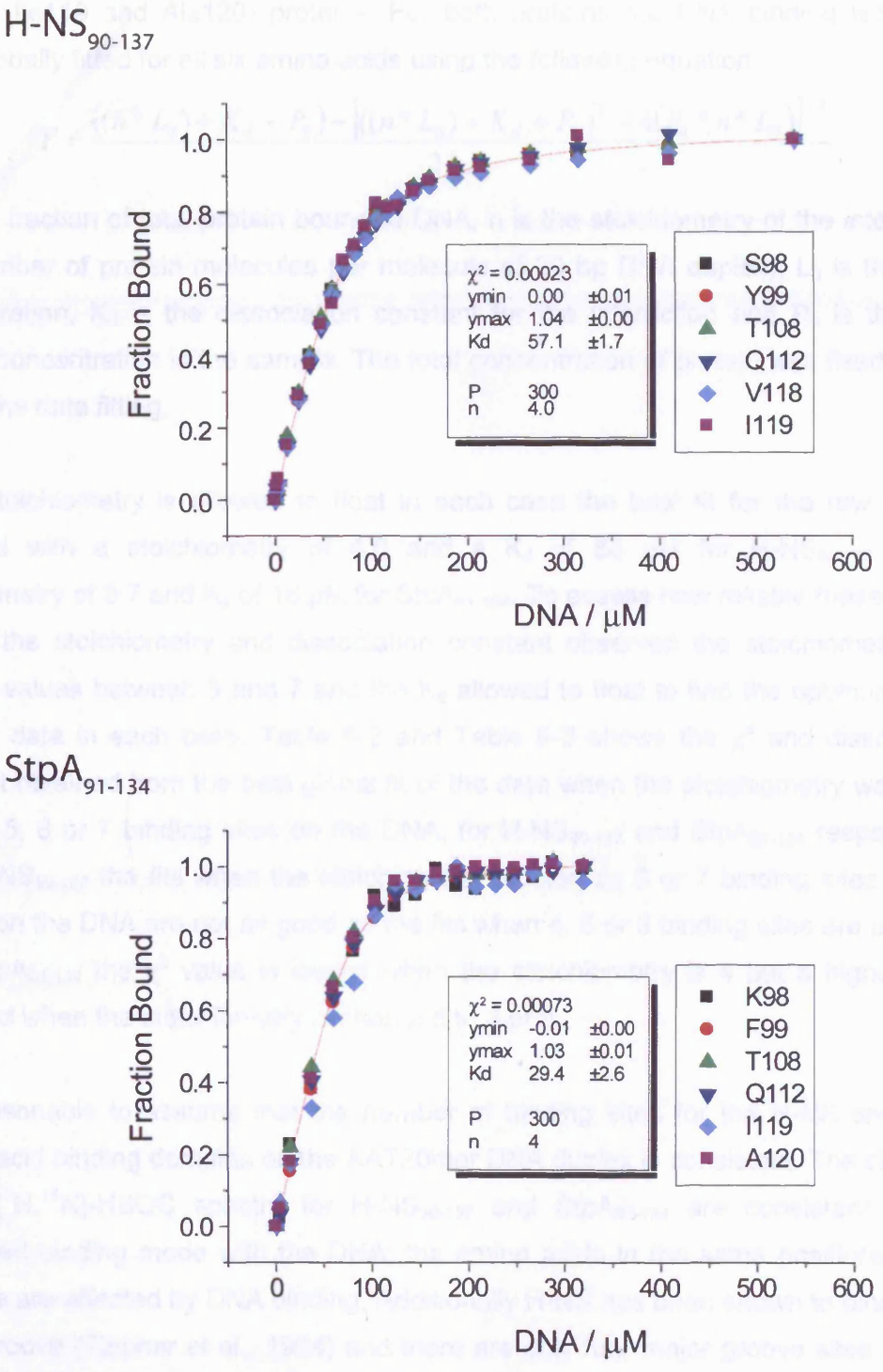


Figure 5-15. DNA binding isotherms for H-NS$_{90-137}$ and StpA$_{91-134}$ using the AAT20mer DNA duplex. The fraction bound was determined from the average chemical shift observed at each point in the DNA titration. Six amino acids were selected as representative of each protein and the change in fraction of protein bound with respect to the DNA concentration was globally fitted by least-squares non-linear regression for all six amino acids. The stoichiometry was fixed at 4.0 and the protein concentration at 300 µM. The values for ymin, ymax and the K_d determined for the best fit are shown along with the corresponding chi^2 value.

Thr108, Gln112, Val118 and Ile119) and $StpA_{91-134}$ (residues Lys98, Phe99, Thr108, Gln112, Ile119 and Ala120) proteins. For both proteins the DNA binding isotherms were globally fitted for all six amino acids using the following equation

$$Y = \frac{((n * L_0) + K_d + P_0) - [((n * L_0) + K_d + P_0)^2 - 4(P_0 * n * L_0)]^{1/2}}{2 * P_0}$$

Y is the fraction of total protein bound to DNA, n is the stoichiometry of the interaction (the number of protein molecules per molecule of 20 bp DNA duplex), L_0 is the DNA concentration, K_d is the dissociation constant for the interaction and P_0 is the total protein concentration in the sample. The total concentration of protein was fixed at 300 μM for the data fitting.

If the stoichiometry is allowed to float in each case the best fit for the raw data is obtained with a stoichiometry of 4.6 and a K_d of 88 μM for $H\text{-}NS_{90-137}$ and a stoichiometry of 3.7 and K_d of 16 μM for $StpA_{91-134}$. To assess how reliable these values are for the stoichiometry and dissociation constant observed the stoichiometry was fixed at values between 3 and 7 and the K_d allowed to float to find the optimum fit for the raw data in each case. Table 5-2 and Table 5-3 shows the χ^2 and dissociation constant obtained from the best global fit of the data when the stoichiometry was fixed at 3, 4, 5, 6 or 7 binding sites on the DNA, for $H\text{-}NS_{90-137}$ and $StpA_{91-134}$ respectively. With $H\text{-}NS_{90-137}$ the fits when the stoichiometry is fixed as 3 or 7 binding sites for the protein on the DNA are not as good as the fits when 4, 5 or 6 binding sites are present. With $StpA_{91-134}$ the χ^2 value is lowest when the stoichiometry is 4 but a higher χ^2 is observed when the stoichiometry is changed to 3 or 5.

It is reasonable to assume that the number of binding sites for the H-NS and StpA nucleic acid binding domains on the AAT20mer DNA duplex is consistent. The changes in the [^{1}H,^{15}N]-HSQC spectra for $H\text{-}NS_{90-137}$ and $StpA_{91-134}$ are consistent with a conserved binding mode with the DNA; the amino acids in the same positions in the structure are affected by DNA binding. Additionally H-NS has been shown to bind in the major groove (Tippner et al., 1994) and there are only four major groove sites in a 20 base pairs AAT20mer DNA duplex so there are four potential binding sites for the nucleic acid binding domain of either H-NS or StpA. It is therefore believed that the stoichiometry of the interaction is 4:1, with four binding sites on the AAT20mer DNA duplex for the H-NS or StpA nucleic acid binding domains.

In Figure 5-12 the global fits for $H\text{-}NS_{90-137}$ and $StpA_{91-134}$ DNA binding are the result of fitting with the stoichiometry fixed to 4.0 (four sites on the DNA where the protein binds). To establish the error in the dissociation constant determined the binding

Table 5-2. Fitting Statistics for H-NS_{91-134} binding to 20 base pairs DNA.

Stoichiometry	χ^2	K_d / μM
3	0.00119	19.4 +/- 2.1
4	0.00023	57.1 +/- 1.7
5	0.00018	106.2 +/- 2.3
6	0.00026	159.8 +/- 3.6
7	0.00035	215.5 +/- 5.1

Table 5-3. Fitting Statistics for $StpA_{91-134}$ binding to 20 base pairs DNA.

Stoichiometry	χ^2	K_d / μM
3	0.00083	4.9 +/- 1.1
4	0.00073	29.4 +/- 2.6
5	0.00112	66.8 +/- 5.4
6	0.00143	109.4 +/- 8.6
7	0.00166	154.4 +/- 12.0

isotherms for the six amino acids selected, for H-NS_{90-137} and $StpA_{91-134}$, were individually fitted to determine the range of values for the dissociation constant that are observed. For H-NS_{90-137} the K_d for the binding of the protein to the 20 base pairs DNA duplex, assuming 4 binding sites for the protein on the DNA, varied from 52.8 μM to 69.1 μM. In the case of $StpA_{91-134}$ the dissociation constant, from the individual fitting of the six selected amino acids, varies from 20.1 μM to 46.6 μM. With the stoichiometry of four binding sites for the nucleic acid binding domain on the DNA, the dissociation constants for the binding of H-NS_{90-137} or $StpA_{91-134}$ to DNA are 57 +/- 12 μM and 29 +/- 17 μM respectively.

5.7 DISCUSSION

StpA and H-NS share 58% sequence identity and have many common functional characteristics. Both proteins have no known DNA sequence specificity but preferentially binding to curved DNA sequences (Sonnenfield et al., 2001; Rimsky et al., 2001; Zuber et al., 1994) and constrain negatively supercoiled DNA (Keatch et al., 2005; Zhang et al., 1996; Cusick and Belfort, 1998). The bridging of DNA by StpA and H-NS to form condensed DNA structures has been visualised by atomic force microscopy (Dame et al., 2000; Dame et al., 2005). Over expression of StpA in a H-NS mutant can suppress many of the H-NS mutant phenotypes (Shi and Bennett, 1994; Sonden and Uhlin, 1996) but StpA does not fully compensate for the absence of H-NS in all cases. The *bgl* operon is regulated by H-NS and StpA is able to partially repress transcription at this promoter in a H-NS mutant (Cusick and Belfort, 1998; Free et al., 2001) but the level of repression does not reach the level observed in the wild type cell even if StpA is over-expressed. Proteolytic degradation of StpA by Lon protease, in the absence of H-NS, results in a lower level of StpA in the cytoplasm in a H-NS mutant compared to the normal level of H-NS in the cell, even if StpA is over expressed (Sonnenfield et al., 2001). The expression of H-NS_{1-93} results in repression of the bgl operon if StpA is present in the cell suggesting that the H-NS oligomerisation domain is able to protect StpA from proteolytic degradation and facilitate the formation of heterooligomeric complexes that can repress *bgl* expression (Free et al., 2001).

The results presented in this chapter describe the first direct comparison of the DNA binding properties of the isolated nucleic acid binding domains of H-NS and StpA. The StpA nucleic acid binding domain has a slightly higher affinity for AT-rich DNA than the H-NS nucleic acid binding domain, which could have important implications for how H-NS, StpA and also heteromeric H-NS:StpA complexes function *in vivo*. A difference in

the NaCl concentration in the buffer can be ruled out as an explanation for the difference in DNA binding affintiy between $StpA_{91-134}$ and $H\text{-}NS_{90-137}$ because the protein samples were exhaustively dialysed at the same time in separate dialysis bags but in the same 2 L beaker to ensure the buffer was identical for both DNA titrations. In addition, the ^{1}H 90° pulse widths calibrated prior to recording the [^{1}H,^{15}N]-HSQC spectra, provide a measure of the ion concentration of the NMR sample as described in section 5.6.3 and show that the buffer contains a comparable level of ions for the DNA titrations of $H\text{-}NS_{90-137}$ and $StpA_{91-134}$. The ^{1}H 90° pulse width increased from 6.875 to 7.175 by the addition of 540 µM AAT20mer double stranded DNA to the ^{15}N-labeled $H\text{-}NS_{90-137}$ NMR sample and the same change in the ^{1}H 90° pulse width was required when 527 µM AAT20mer double stranded DNA was added to the ^{15}N-labeled $StpA_{91-134}$ NMR sample. Both the DNA gel shift assay and the NMR DNA titrations show that the nucleic acid binding domain of StpA binds with higher affinity to the AAT20mer double stranded DNA molecule compared to the nucleic acid binding domain of H-NS.

The affinities of $H\text{-}NS_{90-137}$ or $StpA_{91-134}$ for the AAT20mer DNA duplex determined using NMR spectroscopy are much stronger than the apparent affinity observed by DNA gel shift assay. The binding affinity determined by NMR was under equilibrium conditions. However, a DNA gel shift assay separates the free DNA from the protein:DNA complex so the complex will dissociate as it passes through the gel so the apparent binding affinity will appear lower than the true binding affinity.

Two loop regions, between residues 90 and 96 and 110 and 117 are involved in DNA binding. The DNA binding regions of *E. coli* StpA and *S. typhimurium* H-NS are broadly consistent with previous observations using the *E. coli* H-NS nucleic acid binding domain (Shindo et al., 1999). Shindo *et al* showed that when a homodimer of *E. coli* $H\text{-}NS_{60-137}$ interacts with a 14 base pairs AT-rich DNA duplex two loop regions, spanning from Ala80 to Lys96 and Thr110 to Ala117 respectively, are perturbed by the addition of DNA. The largest change in chemical shift observed upon DNA binding was found in the loop corresponding to residues Thr110 to Ala117 (Shindo et al., 1999) in agreement with the results presented in this chapter. The DNA binding loop between residues 110 and 117 is highly conserved amongst H-NS and StpA homologues (Dorman et al., 1999; Tendeng and Bertin, 2003) whereas residues 80 to 90 are less well conserved and belong to a flexible linker between the oligomerisation domain and the DNA binding domain. Point mutations between Arg90 and Ala95 or Thr110 and Ile119 result in *E. coli* H-NS mutants that are less able to repress the proV promoter *in vivo* compared with wild type H-NS (Ueguchi et al., 1996; Spurio et al., 1997; Williams et al., 1996).

The nucleic acid binding domains of H-NS and StpA have the same secondary structure elements, consisting of two β-strands, an α-helix and a 3_{10}-helix as shown in Figure 5-13. The angle between the β-hairpin and the main α-helix in the *E. coli* $StpA_{91-134}$ structure is greater than the angle observed in the *E. coli* $H\text{-}NS_{91-137}$ structure which could provide an explanation for the increased affinity for DNA of the StpA nucleic acid binding domain. However, the difference in the structures may be the result of variation due to the way the structures were determined. The structure of *E. coli* $StpA_{91-134}$ was solved with significantly more distance restraints and therefore could represent a more accurate determination of the protein fold for these nucleic acid binding domains; 66 long range NOE's and 26 medium range NOE's were used to calculate the structure of *E. coli* $H\text{-}NS_{91-137}$ whereas 163 long range NOE's and 324 medium range NOE's were used to determine the structure of $StpA_{91-134}$. The sixteen lowest energy structures determined for *E. coli* $H\text{-}NS_{91-137}$ show greater variation in the position of the secondary structure elements (Shindo et al., 1995). The root-mean-squared-deviations for the backbone atoms between Tyr97 and Phe133 of *E. coli* $H\text{-}NS_{91-137}$ and Tyr97 and Phe132 of *E. coli* $StpA_{91-134}$ are 2.3 Å and 0.81 Å respectively for the 16 lowest energy model structures.

Alternatively the difference in DNA binding affinity between the *E. coli* H-NS and *S. typhimurium* StpA nucleic acid binding domains could be due to one or more of the fifteen amino acid differences between the two proteins. In particular there are two basic residues, Lys98 and Lys117, which are present in $StpA_{91-134}$ but are replaced by serine and alanine respectively in $H\text{-}NS_{90-137}$. Lys98 and Lys117 are conserved in StpA homologues from *S. typhimurium* and *S. flexneri* (Figure 1-5). The backbone amide cross peak of Lys98 in the $[^1H,^{15}N]$-HSQC spectrum of $StpA_{91-134}$ is only slightly affected by the addition of DNA whereas residue 117 is strongly perturbed upon DNA binding in both the $H\text{-}NS_{90-137}$ and $StpA_{91-134}$ DNA titrations by NMR spectroscopy. Lys117 is located at the C-terminal end of the main DNA binding loop (loop 3 in Figure 5-13) at the beginning of the α-helix. The Lys117 in the StpA nucleic acid binding domain could provide an additional contact with the DNA phosphate backbone relative to the H-NS homologue. It would be interesting to test if a H-NS A117K mutant or a StpA K117A mutant has a different affinity for DNA relative to the wild type proteins.

The increased affinity of StpA for DNA relative to its H-NS homologue could play an important role in the adaptation of enteric bacteria to changes in their environment. The *stpA* gene is up regulated in response to temperature and osmotic stress. An approximately two fold increase in the level of *stpA* mRNA is observed at 37°C

compared to 26°C whereas the concentration of *hns* mRNA did not change in response to this temperature change (Sonden and Uhlin, 1996).

The affinity of H-NS for DNA decreases as the temperature is increased from 20°C to 37°C (Ono et al., 2005; Amit et al., 2003). One role for StpA could be to repress gene expression at promoters that are normally repressed by H-NS at 20°C but where H-NS is no longer able to bind at 37°C. Those genes not required for the response to the elevated temperature could be repressed by StpA because it has a higher affinity for DNA to prevent the change in temperature leading to uncontrolled gene expression. StpA may have a similar role in controlling the adaptation of the enteric bacterial cell to osmotic stress. Increased levels of StpA have been identified in response to higher salt concentrations in the bacterial growth media. In conditions of osmotic stress the concentration of ions inside the cytoplasm is increased to prevent dehydration of the cell. Under these conditions the affinity of H-NS for DNA may be reduced. Force extension measurements of DNA in the presence and absence of H-NS showed that H-NS binding to DNA is reduced by a change in buffer conditions from 50 mM KCl to 200 mM KCl (Amit et al., 2003). The elevated levels of StpA observed in *E. coli* cells under osmotic stress may compensate for the disruption to H-NS mediated repression of transcription by creating StpA complexes and StpA:H-NS heteromeric complexes that have an enhanced affinity for DNA relative to H-NS homooligomers.

6 CONCLUSIONS

The biophysical characteristics of both the oligomerisation domain and the nucleic acid binding domain of H-NS have been investigated in this thesis. The results presented establish for the first time the extent of the oligomerisation domain and determine the residues required for the formation of tetramers and higher order oligomeric species. Whilst H-NS_{1-74} C21S is unable to form tetramers the addition of residues Leu75 to Lys83 produces a stable protein construct that is competent in the formation high order oligomeric species. The DNA binding domain and residues within the flexible linker between the two structured domains are not necessary for the self-association of H-NS.

The H-NS_{1-74} C21S construct, whilst being unable to form tetramers, provides a useful model system to assess the coiled-coil dimerisation interaction. This construct contains the complete predicted coiled-coil region of the protein (residues Leu23 to Arg62) but does not contain the putative fourth α-helix found between residues Pro72 and Ala82 which is essential for the high order oligomerisation of H-NS. The H-NS_{1-74} C21S protein has a melting point of approximately 37°C at 30 μM or higher monomer protein concentration. At lower concentrations of H-NS_{1-74} C21S the melting temperature is concentration dependent showing that the protein is in equilibrium between the unfolded monomer and the folded homodimer at concentrations below 30 μM. These results suggest a possible mechanism for H-NS to function as a temperature sensor in enterobacteria. 408 temperature sensitive genes in *S. typhimurium* have been shown to be either directly or indirectly regulated by H-NS (Ono et al., 2005). The results show that the coiled-coil interaction is disrupted by an increase in temperature resulting in an increase in the proportion of H-NS monomer present when the temperature rises from 20°C to 37°C. These results are in agreement with SEC experiments, which show a decrease in the size of the H-NS_{1-90} protein at the temperature is increased from 17°C to 45°C (Ono et al., 2005). However, previous size exclusion chromatography experiments and suggested that an increase in temperature results in an increase in the size of the average H-NS oligomer (Ceschini et al., 2000). The difference between these SEC results could be due to the different resin used in the experiments which may have different tolerances to temperature change. CD provides a direct measurement of the melting point of the protein so the melting point determined is not affected by systematic errors which could arise in the SEC experiments.

The binding of H-NS to sonicated lambda phage DNA was measured using isothermal titration calorimetry. The change in enthalpy for the interaction with respect to temperature reveals a transition in the structure of H-NS at 35°C which can be explained by an exposure of hydrophobic surface area (Ono et al., 2005). It was proposed that the exposure of hydrophobic surface area could be explained by a transition from a parallel coiled-coil configuration to an anti-parallel coiled-coil H-NS homodimer. The melting of the H-NS coiled-coil to produce unfolded H-NS monomers with a melting point of 37°C provides an alternative explanation for the observed exposure of hydrophobic surface area. Interestingly at 30 μM the $StpA_{1-65}$ protein sample has a melting point of 69°C, measured by CD (unpublished result produced by Dr S. Ono in this laboratory) suggesting that StpA has a more temperature stable coiled-coil than H-NS.

The formation of tetramers and higher order oligomers is important for the function of H-NS and is mediated by the N-terminal 83 amino acids in the protein. The high order oligomerisation of H-NS is disrupted by low ionic strength conditions but when the sodium perchlorate concentration was varied from 100 mM to 300 mM no change in the average size of H-NS was observed. The mutation of Arg15 and Glu73 partially disrupts the formation of tetramers and higher order oligomeric species demonstrating the importance of the second α-helix of H-NS and the putative fourth α-helix on the formation of high order oligomers. Interestingly the $H\text{-}NS_{FL}$ R12E/C21S and $H\text{-}NS_{1-83}$ R12A/C21S mutants have a higher potential for the formation of high order oligomeric species relative to the proteins when Arg12 is present. The role of this amino acid in reducing the affinity of the self-association interactions is unclear. Arg12 has also been identified as important for the interaction of Hha with H-NS. It has been suggested that Hha mimics the structure of H-NS (Madrid et al., 2007b) but the Hha monomer, whose structure has been solved by NMR, undergoes a change in structure to a more open configuration upon binding to $H\text{-}NS_{1-65}$ (Garcia et al., 2005) so it is not known what the tertiary structure of Hha is when bound to H-NS.

The $H\text{-}NS_{FL}$ R15E C21S mutant is disrupted in its potential to form homotetramers. This amino acid forms a stabilising salt bridge in the published NMR structures of *S. typhimurium* $H\text{-}NS_{2-58}$ C21S and *E. coli* $H\text{-}NS_{2-47}$ so the mutation of this amino acid could affect the position of the second α-helix in the H-NS protein. Whilst the disruption to the high order oligomerisation of H-NS by the R15E mutation, provides an alternative explanation for why the H-NS R12E/R15A mutant has reduced affinity for DNA to the interpretation that this arginine is involved in DNA binding directly (Bloch et al., 2003;

Garcia et al., 2005; Paytubi et al., 2004), this cannot be taken as evidence against the antiparallel coiled-coil structure of H-NS.

The structure of a complete H-NS homodimer needs to be determined to establish the correct configuration of the coiled-coil. Low ionic strength conditions have been used to establish conditions suitable for the crystallisation of the oligomerisation domain of H-NS. Although protein crystals of both H-NS_{1-71} C21S and H-NS_{1-83} C21S could be obtained suitable conditions were not identified that produce crystals of sufficient resolution to determine the protein structure. It is not possible to determine whether or not the crystallisation conditions established represent a significant step towards the determination of the structure of the oligomerisation domain. However, there are a couple of avenues that could be followed to use these crystallisation conditions as a starting point to determine the protein structure of H-NS. Firstly alternative additives or crystal soaks could be attempted using crystals produced under the conditions identified. Another possible approach would be to place the H-NS_{1-83} C21S protein crystals in an NMR tube and use solid state NMR techniques to determine the protein structure. Having the protein in crystalline form will block the formation of the high order oligomeric species of H-NS that prevent the use of solution state NMR for the determination of the complete oligomerisation domain structure.

The DNA binding interaction of the H-NS and StpA nucleic acid binding domains have also been investigated by NMR. Despite the similarity in the DNA binding region the titration of DNA into separate NMR samples of *S. typhimurium* H-NS_{90-137} and *E. coli* $StpA_{91-134}$ reveals that the StpA nucleic acid binding domain has a two-fold higher binding affinity for DNA relative to the H-NS nucleic acid binding domain. The difference in affinity for DNA suggests a potential role for StpA in modulating the repression of gene expression in the cell by H-NS. The auto-regulation of the *hns* promoter is linked to DNA synthesis so that a relatively constant ratio of H-NS to DNA is present in the cell (Free and Dorman, 1995). On the other hand, the *stpA* gene expression is strongly repressed by H-NS but is responsive to changes in environmental conditions (Free and Dorman, 1997). The formation of heteromeric complexes between H-NS and StpA will result in the formation of a protein complex with unique DNA binding properties compared to a complex formed from H-NS monomers alone. Disruption to H-NS mediated gene repression, by an increase in temperature for example, will result in greater expression of the *stpA* gene and therefore the H-NS homomeric complexes in the enteric bacteria will become increasingly heteromeric complexes formed from both H-NS and StpA. These H-

NS:StpA complexes will bind more tightly to the DNA than the equivalent size of H-NS homooligomer.

7 REFERENCES

Ackers,G.K. (1967). A New Calibration Procedure for Gel Filtration Columns. J Biol. Chem. *242*, 3237-3238.

Afflerbach,H., Schroder,O., and Wagner,R. (1999). Conformational changes of the upstream DNA mediated by H-NS and FIS regulate *E. coli* RrnB P1 promoter activity. J. Mol. Biol. *286*, 339-353.

Ali,B.M., Amit,R., Braslavsky,I., Oppenheim,A.B., Gileadi,O., and Stavans,J. (2001). Compaction of single DNA molecules induced by binding of integration host factor (IHF). Proc. Natl. Acad. Sci. U. S. A *98*, 10658-10663.

Amit,R., Oppenheim,A.B., and Stavans,J. (2003). Increased bending rigidity of single DNA molecules by H-NS, a temperature and osmolarity sensor. Biophys. J. *84*, 2467-2473.

Arnqvist,A., Olsen,A., and Normark,S. (1994). Sigma S-dependent growth-phase induction of the csgBA promoter in *Escherichia coli* can be achieved *in vivo* by sigma 70 in the absence of the nucleoid-associated protein H-NS. Mol. Microbiol. *13*, 1021-1032.

Atlung,T. and Ingmer,H. (1997). H-NS: a modulator of environmentally regulated gene expression. Mol. Microbiol. *24*, 7-17.

Azam,T.A., Hiraga,S., and Ishihama,A. (2000). Two types of localization of the DNA-binding proteins within the *Escherichia coli* nucleoid. Genes Cells *5*, 613-626.

Azam,T.A., Iwata,A., Nishimura,A., Ueda,S., and Ishihama,A. (1999). Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. J Bacteriol. *181*, 6361-6370.

Badaut,C., Williams,R., Arluison,V., Bouffartigues,E., Robert,B., Buc,H., and Rimsky,S. (2002). The degree of oligomerization of the H-NS nucleoid structuring protein is related to specific binding to DNA. J. Biol. Chem. *277*, 41657-41666.

Bang,I.S., Audia,J.P., Park,Y.K., and Foster,J.W. (2002). Autoinduction of the *ompR* response regulator by acid shock and control of the *Salmonella enterica* acid tolerance response. Mol. Microbiol. *44*, 1235-1250.

Bax,A., Clore,G.M., and Gronenborn,A.M. (1990). Detection of nuclear Overhauser effects between degenerate amide proton resonances by heteronuclear three-dimensional NMR spectroscopy. J. Am. Chem. Soc. *112*, 9020-9021.

Benesch,J.L. and Robinson,C.V. (2006). Mass spectrometry of macromolecular assemblies: preservation and dissociation. Curr. Opin. Struct. Biol. *16*, 245-251.

Bergfors,T. (1999). Protein Crystallization. Techniques, Strategies, and Tips. A Laboratory Manual. Methods Mol. Biol. *363*, 131-151.

Bertin,P., Benhabiles,N., Krin,E., Laurent-Winter,C., Tendeng,C., Turlin,E., Thomas,A., Danchin,A., and Brasseur,R. (1999). The structural and functional organization of H-NS-like proteins is evolutionarily conserved in gram-negative bacteria. Mol. Microbiol. *31*, 319-329.

Bertin,P., Hommais,F., Krin,E., Soutourina,O., Tendeng,C., Derzelle,S., and Danchin,A. (2001). H-NS and H-NS-like proteins in Gram-negative bacteria and their multiple role in the regulation of bacterial metabolism. Biochimie *83*, 235-241.

Bjellqvist,B., Hughes,G.J., Pasquali,C., Paquet,N., Ravier,F., Sanchez,J.C., Frutiger,S., and Hochstrasser,D. (1993). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis *14*, 1023-1031.

Blattner,F.R., Plunkett,G., III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B., and Shao,Y. (1997). The complete genome sequence of *Escherichia coli* K-12. Science *277*, 1453-1474.

Bloch,V., Yang,Y., Margeat,E., Chavanieu,A., Auge,M.T., Robert,B., Arold,S., Rimsky,S., and Kochoyan,M. (2003). The H-NS dimerization domain defines a new fold contributing to DNA recognition. Nat. Struct. Biol. *10*, 212-218.

Bolbach,G. (2005). Matrix-assisted laser desorption/ionization analysis of non-covalent complexes: fundamentals and applications. Curr. Pharm. Des *11*, 2535-2557.

Bouffartigues,E., Buckle,M., Badaut,C., Travers,A., and Rimsky,S. (2007). H-NS cooperative binding to high-affinity sites in a regulatory element results in transcriptional silencing. Nat. Struct. Mol. Biol. *14*, 441-448.

Browning,D.F., Cole,J.A., and Busby,S.J. (2000). Suppression of FNR-dependent transcription activation at the *Escherichia coli nir* promoter by Fis, IHF and H-NS: modulation of transcription initiation by a complex nucleo-protein assembly. Mol. Microbiol. *37*, 1258-1269.

Cabrera,J.E. and Jin,D.J. (2003). The distribution of RNA polymerase in *Escherichia coli* is dynamic and sensitive to environmental cues. Mol. Microbiol. *50*, 1493-1505.

Calvo,J.M. and Matthews,R.G. (1994). The leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. Microbiol. Rev. *58*, 466-490.

Case,R.B., Chang,Y.P., Smith,S.B., Gore,J., Cozzarelli,N.R., and Bustamante,C. (2004). The bacterial condensin MukBEF compacts DNA into a repetitive, stable structure. Science *305*, 222-227.

Cerdan,R., Bloch,V., Yang,Y., Bertin,P., Dumas,C., Rimsky,S., Kochoyan,M., and Arold,S.T. (2003). Crystal structure of the N-terminal dimerisation domain of VicH, the H-NS-like protein of *Vibrio cholerae*. J. Mol. Biol. *334*, 179-185.

Ceschini,S., Lupidi,G., Coletta,M., Pon,C.L., Fioretti,E., and Angeletti,M. (2000). Multimeric self-assembly equilibria involving the histone-like protein H-NS. A thermodynamic study. J. Biol. Chem. *275*, 729-734.

Cheng,K.Y., Lowe,E.D., Sinclair,J., Nigg,E.A., and Johnson,L.N. (2003). The crystal structure of the human polo-like kinase-1 polo box domain and its phospho-peptide complex. EMBO J *22*, 5757-5768.

Chothia,C. and Lesk,A.M. (1986). The relation between the divergence of sequence and structure in proteins. EMBO J *5*, 823-826.

Cudney,R., Patel,S., Weisgraber,K., Newhouse,Y., and McPherson,A. (1994). Screening and optimization strategies for macromolecular crystal growth. Acta Crystallogr. D. Biol. Crystallogr. *50*, 414-423.

Cuff,J.A., Clamp,M.E., Siddiqui,A.S., Finlay,M., and Barton,G.J. (1998). JPred: a consensus secondary structure prediction server. Bioinformatics. *14*, 892-893.

Cusick,M.E. and Belfort,M. (1998). Domain structure and RNA annealing activity of the *Escherichia coli* regulatory protein StpA. Mol. Microbiol. *28*, 847-857.

D'Arcy,A. (1994). Crystallizing proteins - a rational approach? Acta Crystallogr. D. Biol. Crystallogr. *50*, 469-471.

Dam,J. and Schuck,P. (2004). Calculating sedimentation coefficient distributions by direct modeling of sedimentation velocity concentration profiles. Methods Enzymol. *384*, 185-212.

Dame,R.T. and Goosen,N. (2002). HU: promoting or counteracting DNA compaction? FEBS Lett. *529*, 151-156.

Dame,R.T., Luijsterburg,M.S., Krin,E., Bertin,P.N., Wagner,R., and Wuite,G.J. (2005). DNA bridging: a property shared among H-NS-like proteins. J. Bacteriol. *187*, 1845-1848.

Dame,R.T., Noom,M.C., and Wuite,G.J. (2006). Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. Nature *444*, 387-390.

Dame,R.T., Wyman,C., and Goosen,N. (2000). H-NS mediated compaction of DNA visualised by atomic force microscopy. Nucleic Acids Res. *28*, 3504-3510.

Dame,R.T., Wyman,C., Wurm,R., Wagner,R., and Goosen,N. (2002). Structural basis for H-NS-mediated trapping of RNA polymerase in the open initiation complex at the rrnB P1. J. Biol. Chem. *277*, 2146-2150.

Defez,R. and De,F.M. (1982). The metabolism of beta-glucosides in *Escherichia coli* K12. Ann. Microbiol. (Paris) *133*, 347-350.

Deighan,P., Beloin,C., and Dorman,C.J. (2003). Three-way interactions among the Sfh, StpA and H-NS nucleoid-structuring proteins of *Shigella flexneri* 2a strain 2457T. Mol. Microbiol. *48*, 1401-1416.

Deighan,P., Free,A., and Dorman,C.J. (2000). A role for the *Escherichia coli* H-NS-like protein StpA in OmpF porin expression through modulation of *micF* RNA stability. Mol. Microbiol. *38*, 126-139.

Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J., and Bax,A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol. NMR *6*, 277-293.

Delano,W.L. (2006). The PyMOL Molecular Graphics System; http://www.pymol.org. http://www. pymol. org.

Delihas,N. and Forst,S. (2001). *MicF*: an antisense RNA gene involved in response of *Escherichia coli* to global stress factors. J. Mol. Biol. *313*, 1-12.

Delorenzi,M. and Speed,T. (2002). An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics. *18*, 617-625.

Deng,S., Stein,R.A., and Higgins,N.P. (2004). Transcription-induced barriers to supercoil diffusion in the *Salmonella typhimurium* chromosome. Proc. Natl. Acad. Sci. U. S. A *101*, 3398-3403.

Denisov,V.P. and Halle,B. (2000). Sequence-specific binding of counterions to B-DNA. Proc. Natl. Acad. Sci. U. S. A *97*, 629-633.

Derewenda,Z.S. (2004). Rational protein crystallization by mutational surface engineering. Structure. *12*, 529-535.

Dersch,P., Schmidt,K., and Bremer,E. (1993). Synthesis of the *Escherichia coli* K-12 nucleoid-associated DNA-binding protein H-NS is subjected to growth-phase control and autoregulation. Mol. Microbiol. *8*, 875-889.

Donato,G.M. and Kawula,T.H. (1999). Phenotypic analysis of random *hns* mutations differentiate DNA-binding activity from properties of fimA promoter inversion modulation and bacterial motility. J. Bacteriol. *181*, 941-948.

Dorman,C.J. (2004). H-NS: a universal regulator for a dynamic genome. Nat. Rev. Microbiol. *2*, 391-400.

Dorman,C.J. (2007a). Probing bacterial nucleoid structure with optical tweezers. Bioessays *29*, 212-216.

Dorman,C.J., Hinton,J.C., and Free,A. (1999). Domain organization and oligomerization among H-NS-like nucleoid-associated proteins in bacteria. Trends Microbiol. *7*, 124-128.

Dorman,C.J. (2007b). H-NS, the genome sentinel. Nat Rev Micro *5*, 157-161.

Drlica,K. (1992). Control of bacterial DNA supercoiling. Molecular Microbiology *6*, 425-433.

Esposito,D., Petrovic,A., Harris,R., Ono,S., Eccleston,J.F., Mbabaali,A., Haq,I., Higgins,C.F., Hinton,J.C., Driscoll,P.C., and Ladbury,J.E. (2002). H-NS oligomerization domain structure reveals the mechanism for high order self-association of the intact protein. J. Mol. Biol. *324*, 841-850.

Eswar,N., John,B., Mirkovic,N., Fiser,A., Ilyin,V.A., Pieper,U., Stuart,A.C., Marti-Renom,M.A., Madhusudhan,M.S., Yerkovich,B., and Sali,A. (2003). Tools for comparative protein structure modeling and analysis. Nucleic Acids Res. *31*, 3375-3380.

Evans,J.N.S. (1996). Biomolecular NMR Spectroscopy. Oxford University Press).

Falconi,M., Brandi,A., La,T.A., Gualerzi,C.O., and Pon,C.L. (1996). Antagonistic involvement of FIS and H-NS proteins in the transcriptional control of *hns* expression. Mol. Microbiol. *19*, 965-975.

Falconi,M., Colonna,B., Prosseda,G., Micheli,G., and Gualerzi,C.O. (1998). Thermoregulation of *Shigella* and *Escherichia coli* EIEC pathogenicity. A temperature-dependent structural transition of DNA modulates accessibility of *virF* promoter to transcriptional repressor H-NS. EMBO J. *17*, 7033-7043.

Falconi,M., Higgins,N.P., Spurio,R., Pon,C.L., and Gualerzi,C.O. (1993). Expression of the gene encoding the major bacterial nucleotide protein H-NS is subject to transcriptional auto-repression. Mol. Microbiol. *10*, 273-282.

Falconi,M., Prosseda,G., Giangrossi,M., Beghetto,E., and Colonna,B. (2001). Involvement of FIS in the H-NS-mediated regulation of *virF* gene of *Shigella* and enteroinvasive *Escherichia coli*. Mol. Microbiol. *42*, 439-452.

Farmer,T.B. and Caprioli,R.M. (1998). Determination of protein-protein interactions by matrix-assisted laser desorption/ionization mass spectrometry. J Mass Spectrom. *33*, 697-704.

Farrow,N.A., Muhandiram,R., Singer,A.U., Pascal,S.M., Kay,C.M., Gish,G., Shoelson,S.E., Pawson,T., Forman-Kay,J.D., and Kay,L.E. (1994). Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by 15N NMR Relaxation. Biochemistry *33*, 5984-6003.

Ferre-D'Amare,A.R. and Burley,S.K. (1994). Use of dynamic light scattering to assess crystallizability of macromolecules and macromolecular assemblies. Structure. *2*, 357-359.

Free,A. and Dorman,C.J. (1995). Coupling of *Escherichia coli hns* mRNA levels to DNA synthesis by autoregulation: implications for growth phase control. Mol. Microbiol. *18*, 101-113.

Free,A. and Dorman,C.J. (1997). The *Escherichia coli stpA* gene is transiently expressed during growth in rich medium and is induced in minimal medium and by stress conditions. J. Bacteriol. *179*, 909-918.

Free,A., Porter,M.E., Deighan,P., and Dorman,C.J. (2001). Requirement for the molecular adapter function of StpA at the *Escherichia coli bgl* promoter depends upon the level of truncated H-NS protein. Mol. Microbiol. *42*, 903-917.

Frenkiel-Krispin,D., Ben-Avraham,I., Englander,J., Shimoni,E., Wolf,S.G., and Minsky,A. (2004). Nucleoid restructuring in stationary-state bacteria. Mol. Microbiol. *51*, 395-405.

Garcia,J., Cordeiro,T.N., Nieto,J.M., Pons,I., Juarez,A., and Pons,M. (2005). Interaction between the bacterial nucleoid associated proteins Hha and H-NS involves a conformational change of Hha. Biochem. J. *388*, 755-762.

Garcia,J., Madrid,C., Juarez,A., and Pons,M. (2006). New roles for key residues in helices H1 and H2 of the *Escherichia coli* H-NS N-terminal domain: H-NS dimer stabilization and Hha binding. J. Mol. Biol. *359*, 679-689.

Gill,S.C. and von Hippel,P.H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem. *182*, 319-326.

Goransson,M., Sonden,B., Nilsson,P., Dagberg,B., Forsman,K., Emanuelsson,K., and Uhlin,B.E. (1990). Transcriptional silencing and thermoregulation of gene expression in *Escherichia coli*. Nature *344*, 682-685.

Grainger,D.C., Hurd,D., Goldberg,M.D., and Busby,S.J. (2006). Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. Nucleic Acids Res. *34*, 4642-4652.

Grant,R.A., Filman,D.J., Finkel,S.E., Kolter,R., and Hogle,J.M. (1998). The crystal structure of Dps, a ferritin homolog that binds and protects DNA. Nat Struct. Biol. *5*, 294-303.

Graumann,P.L. (2001). SMC proteins in bacteria: condensation motors for chromosome segregation? Biochimie *83*, 53-59.

Greenfield,N.J. (1996). Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. Anal. Biochem. *235*, 1-10.

Groisman,E.A. and Ochman,H. (1997). How *Salmonella* became a pathogen. Trends Microbiol. *5*, 343-349.

Gruber,M., Soding,J., and Lupas,A.N. (2006). Comparative analysis of coiled-coil prediction methods. J. Struct. Biol. *155*, 140-145.

Haack,K.R., Robinson,C.L., Miller,K.J., Fowlkes,J.W., and Mellies,J.L. (2003). Interaction of Ler at the LEE5 (tir) operon of enteropathogenic *Escherichia coli*. Infect. Immun. *71*, 384-392.

Hall,T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. *Ser. 41*, 95-98.

Hare,S., Bayliss,R., Baron,C., and Waksman,G. (2006). A large domain swap in the VirB11 ATPase of *Brucella suis* leaves the hexameric assembly intact. Journal of Molecular Biology *360*, 56-66.

Heck,A.J. and Van Den Heuvel,R.H. (2004). Investigation of intact protein complexes by mass spectrometry. Mass Spectrom. Rev. *23*, 368-389.

Hendrickson,W.A., Smith,J.L., and Sheriff,S. (1985). Direct phase determination based on anomalous scattering. Methods Enzymol. *115*, 41-55.

Heras,B., Edeling,M.A., Byriel,K.A., Jones,A., Raina,S., and Martin,J.L. (2003). Dehydration converts DsbG crystal diffraction from low to high resolution. Structure. *11*, 139-145.

Heras,B. and Martin,J.L. (2005). Post-crystallization treatments for improving diffraction quality of protein crystals. Acta Crystallogr. D. Biol. Crystallogr. *61*, 1173-1180.

Higgins,N.P., Yang,X., Fu,Q., and Roth,J.R. (1996). Surveying a supercoil domain by using the gamma delta resolution system in *Salmonella typhimurium*. J Bacteriol. *178*, 2825-2835.

Hommais,F., Krin,E., Laurent-Winter,C., Soutourina,O., Malpertuy,A., Le Caer,J.P., Danchin,A., and Bertin,P. (2001). Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. Mol. Microbiol. *40*, 20-36.

Hud,N.V. and Polak,M. (2001). DNA-cation interactions: The major and minor grooves are flexible ionophores. Curr. Opin. Struct. Biol. *11*, 293-301.

Hurme,R., Berndt,K.D., Normark,S.J., and Rhen,M. (1997). A proteinaceous gene regulatory thermometer in *Salmonella*. Cell *90*, 55-64.

Jacquet,M., Cukier-Kahn,R., Pla,J., and Gros,F. (1971). A thermostable protein factor acting on *in vitro* DNA transcription. Biochem. Biophys. Res. Commun. *45*, 1597-1607.

Jancarik,J. and Kim,S.H. (1994). Sparse matrix sampling: a screening metod for crystallization of proteins. Acta Crystallogr. D. Biol. Crystallogr. *24*, 409-411.

Johansson,J. and Uhlin,B.E. (1999). Differential protease-mediated turnover of H-NS and StpA revealed by a mutation altering protein stability and stationary-phase survival of *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A *96*, 10776-10781.

Johnson,R.C., Bruist,M.F., and Simon,M.I. (1986). Host protein requirements for *in vitro* site-specific DNA inversion. Cell *46*, 531-539.

Jordi,B.J., Fielder,A.E., Burns,C.M., Hinton,J.C., Dover,N., Ussery,D.W., and Higgins,C.F. (1997). DNA binding is not sufficient for H-NS-mediated repression of *proU* expression. J. Biol. Chem. *272*, 12083-12090.

Jordi,B.J. and Higgins,C.F. (2000). The downstream regulatory element of the proU operon of Salmonella typhimurium inhibits open complex formation by RNA polymerase at a distance. J. Biol. Chem. *275*, 12123-12128.

Kar,S., Edgar,R., and Adhya,S. (2005). Nucleoid remodeling by an altered HU protein: Reorganization of the transcription program. Proc. Natl. Acad. Sci. U. S. A *102*, 16397-16402.

Karem,K. and Foster,J.W. (1993). The influence of DNA topology on the environmental regulation of a pH-regulated locus in *Salmonella typhimurium*. Mol. Microbiol. *10*, 75-86.

Kay,L.E., Keifer,P., and Saarinen,T. (1992). Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J. Am. Chem. Soc. *114*, 10663-10665.

Keatch,S.A., Leonard,P.G., Ladbury,J.E., and Dryden,D.T. (2005). StpA protein from *Escherichia coli* condenses supercoiled DNA in preference to linear DNA and protects it from digestion by DNase I and EcoKI. Nucleic Acids Res. *33*, 6540-6546.

Kiefersauer,R., Stetefeld,J., Gomis-Ruth,F.X., Romao,M.J., Lottspeich,F., and Huber,R. (1996). Protein-Crystal Density by Volume Measurement and Amino-Acid Analysis. Journal of Applied Crystallography *29*, 311-317.

Kiefersauer,R., Than,M.E., Dobbek,H., Gremer,L., Melero,M., Strobl,S., Dias,J.M., Soulimane,T., and Huber,R. (2000). A novel free-mounting system for protein crystals: transformation and improvement of diffraction power by accurately controlled humidity changes. Journal of Applied Crystallography *33*, 1223-1230.

Kimber,M.S., Vallee,F., Houston,S., Necakov,A., Skarina,T., Evdokimova,E., Beasley,S., Christendat,D., Savchenko,A., Arrowsmith,C.H., Vedadi,M., Gerstein,M., and Edwards,A.M. (2003). Data mining crystallization databases: knowledge-based approaches to optimize protein crystal screens. Proteins *51*, 562-568.

Kiselar,J.G. and Downard,K.M. (2000). Preservation and detection of specific antibody--peptide complexes by matrix-assisted laser desorption ionization mass spectrometry. J Am. Soc. Mass Spectrom. *11*, 746-750.

Kostrewa,D., Granzin,J., Stock,D., Choe,H.W., Labahn,J., and Saenger,W. (1992). Crystal structure of the factor for inversion stimulation FIS at 2.0 A resolution. J Mol. Biol. *226*, 209-226.

Kraulis,P.J. (1989). ANSIG: A program for the Assignment of Protein 1H 2D NMR Spectra by Interactive Computer Graphics. Journal of Magnetic Resonance *84*, 627-633.

Kraulis,P.J., Domaille,P.J., Campbell-Burk,S.L., Van,A.T., and Laue,E.D. (1994). Solution structure and dynamics of ras p21.GDP determined by heteronuclear three- and four-dimensional NMR spectroscopy. Biochemistry *33*, 3515-3531.

Kuo,A., Bowler,M.W., Zimmer,J., Antcliff,J.F., and Doyle,D.A. (2003). Increasing the diffraction limit and internal order of a membrane protein crystal by dehydration. J Struct. Biol. *141*, 97-102.

Lang,B., Blot,N., Bouffartigues,E., Buckle,M., Geertz,M., Gualerzi,C.O., Mavathur,R., Muskhelishvili,G., Pon,C.L., Rimsky,S., Stella,S., Babu,M.M., and Travers,A. (2007). High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. Nucleic Acids Res *35*, 6330-6337.

Laskowski,R.A., MacArthur,M.W., Moss,D.S., and Thornton,J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography *26*, 283-291.

Lattman,E. (1985). Diffraction methods for biological macromolecules. Use of the rotation and translation functions. Methods Enzymol. *115*, 55-77.

Laue,T.M., Shah,B.D., Ridgeway,T.M., and Pelletier,S.L. (1992). Analytical Ultracentrifugation in Biochemistry and Polymer Science 90-125.

Lease,R.A. and Belfort,M. (2000). A trans-acting RNA as a control switch in *Escherichia coli*: DsrA modulates function by forming alternative structures. Proc. Natl. Acad. Sci. U. S. A *97*, 9919-9924.

Lease,R.A., Cusick,M.E., and Belfort,M. (1998). Riboregulation in *Escherichia coli*: DsrA RNA acts by RNA:RNA interactions at multiple loci. Proc. Natl. Acad. Sci. U. S. A *95*, 12456-12461.

Lebowitz,J., Lewis,M.S., and Schuck,P. (2002). Modern analytical ultracentrifugation in protein science: a tutorial review. Protein Sci. *11*, 2067-2079.

Leirmo,S., Harrison,C., Cayley,D.S., Burgess,R.R., and Record,M.T., Jr. (1987). Replacement of potassium chloride by potassium glutamate dramatically enhances protein-DNA interactions *in vitro*. Biochemistry *26*, 2095-2101.

Lerman,L.S. (1971). A transition to a compact form of DNA in polymer solutions. Proc. Natl. Acad. Sci. U. S. A *68*, 1886-1890.

Lin,S.H. and Lee,J.C. (2003). Determinants of DNA bending in the DNA-cyclic AMP receptor protein complexes in Escherichia coli. Biochemistry *42*, 4809-4818.

Lorber,B., Sauter,C., Ng,J.D., Zhu,D.W., Giege,R., Vidal,O., Robert,M.C., and Capelle,B. (1999). Characterization of protein and virus crystals by quasi-planar wave X-ray topography: a comparison between crystals grown in solution and in agarose gel. Journal of Crystal Growth *204*, 357-368.

Lucchini,S., Rowley,G., Goldberg,M.D., Hurd,D., Harrison,M., and Hinton,J.C. (2006). H-NS mediates the silencing of laterally acquired genes in bacteria. PLoS. Pathog. *2*, e81.

Lucht,J.M., Dersch,P., Kempf,B., and Bremer,E. (1994). Interactions of the nucleoid-associated DNA-binding protein H-NS with the regulatory region of the osmotically controlled proU operon of *Escherichia coli*. J. Biol. Chem. *269*, 6578.

Luijsterburg,M.S., Noom,M.C., Wuite,G.J., and Dame,R.T. (2006). The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. J Struct. Biol. *156*, 262-272.

Lusty,C.J. (1999). A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography. Journal of Applied Crystallography *32*, 106-112.

Lynch,T.W., Read,E.K., Mattis,A.N., Gardner,J.F., and Rice,P.A. (2003). Integration host factor: putting a twist on protein-DNA recognition. J Mol. Biol. *330*, 493-502.

Madrid,C., Balsalobre,C., Garcia,J., and Juarez,A. (2007b). The novel Hha/YmoA family of nucleoid-associated proteins: use of structural mimicry to modulate the activity of the H-NS family of proteins. Molecular Microbiology *63*, 7-14.

Madrid,C., Balsalobre,C., Garcia,J., and Juarez,A. (2007a). The novel Hha/YmoA family of nucleoid-associated proteins: use of structural mimicry to modulate the activity of the H-NS family of proteins. Molecular Microbiology *63*, 7-14.

Marion,D., Ikura,M., Tschudin,R., and Bax,A. (1989). Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. J. Magn. Reson. *85*, 393-399.

Marschall,C., Labrousse,V., Kreimer,M., Weichart,D., Kolb,A., and Hengge-Aronis,R. (1998). Molecular analysis of the regulation of csiD, a carbon starvation-inducible gene in *Escherichia coli* that is exclusively dependent on sigma s and requires activation by cAMP-CRP. J. Mol. Biol. *276*, 339-353.

Mayer,O., Waldsich,C., Grossberger,R., and Schroeder,R. (2002). Folding of the td pre-RNA with the help of the RNA chaperone StpA. Biochem. Soc. Trans. *30*, 1175-1180.

McCoy,A.J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr. D Biol. Crystallogr. *63*, 32-41.

McGovern,V., Higgins,N.P., Chiz,R.S., and Jaworski,A. (1994). H-NS over-expression induces an artificial stationary phase by silencing global transcription. Biochimie *76*, 1019-1029.

McGuffin,L.J., Bryson,K., and Jones,D.T. (2000). The PSIPRED protein structure prediction server. Bioinformatics. *16*, 404-405.

Muller,C.M., Dobrindt,U., Nagy,G., Emody,L., Uhlin,B.E., and Hacker,J. (2006). Role of Histone-Like Proteins H-NS and StpA in Expression of Virulence Determinants of Uropathogenic *Escherichia coli*. J. Bacteriol. *188*, 5428-5438.

Murtin,C., Engelhorn,M., Geiselmann,J., and Boccard,F. (1998). A quantitative UV laser footprinting analysis of the interaction of IHF with specific binding sites: re-evaluation of the effective concentration of IHF in the cell. J Mol. Biol. *284*, 949-961.

Musacchio,A., Saraste,M., and Wilmanns,M. (1994). High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides. Nat. Struct. Biol. *1*, 546-551.

Naik,R.R., Kirkpatrick,S.M., and Stone,M.O. (2001). The thermostability of an alpha-helical coiled-coil protein and its potential use in sensor applications. Biosens. Bioelectron. *16*, 1051-1057.

Navarre,W.W., Porwollik,S., Wang,Y., McClelland,M., Rosen,H., Libby,S.J., and Fang,F.C. (2006). Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. Science *313*, 236-238.

Neidhardt,F.C., Bloch,P.L., and Smith,D.F. (1974). Culture medium for enterobacteria. J. Bacteriol. *119*, 736-747.

Nieto,J.M., Madrid,C., Miquelay,E., Parra,J.L., Rodriguez,S., and Juarez,A. (2002). Evidence for direct protein-protein interaction between members of the enterobacterial Hha/YmoA and H-NS families of proteins. J. Bacteriol. *184*, 629-635.

O'Byrne,C.P. and Dorman,C.J. (1994). Transcription of the *Salmonella typhimurium* spv virulence locus is regulated negatively by the nucleoid-associated protein H-NS. FEMS Microbiol. Lett. *121*, 99-105.

Oberto,J., Drlica,K., and Rouviere-Yaniv,J. (1994). Histones, HMG, HU, IHF: Meme combat. Biochimie *76*, 901-908.

Ochman,H., Lawrence,J.G., and Groisman,E.A. (2000). Lateral gene transfer and the nature of bacterial innovation. Nature *405*, 299-304.

Ohniwa,R.L., Morikawa,K., Kim,J., Ohta,T., Ishihama,A., Wada,C., and Takeyasu,K. (2006). Dynamic state of DNA topology is essential for genome condensation in bacteria. EMBO J *25*, 5591-5602.

Ohyama,T. (2001). Intrinsic DNA bends: an organizer of local chromatin structure for transcription. Bioessays *23*, 708-715.

Ono,S., Goldberg,M.D., Olsson,T., Esposito,D., Hinton,J.C., and Ladbury,J.E. (2005). H-NS is a part of a thermally controlled mechanism for bacterial gene regulation. Biochem. J. *391*, 203-213.

Oshima,T., Ishikawa,S., Kurokawa,K., Aiba,H., and Ogasawara,N. (2006). *Escherichia coli* histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase. DNA Res *13*, 141-153.

Ouali,M. and King,R.D. (2000). Cascaded multiple classifiers for secondary structure prediction. Protein Sci. *9*, 1162-1176.

Owen-Hughes,T.A., Pavitt,G.D., Santos,D.S., Sidebotham,J.M., Hulton,C.S., Hinton,J.C., and Higgins,C.F. (1992). The chromatin-associated protein H-NS interacts with curved DNA to influence DNA topology and gene expression. Cell *71*, 255-265.

Pan,C.Q., Finkel,S.E., Cramton,S.E., Feng,J.A., Sigman,D.S., and Johnson,R.C. (1996). Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts. J Mol. Biol. *264*, 675-695.

Pantazatos,D., Kim,J.S., Klock,H.E., Stevens,R.C., Wilson,I.A., Lesley,S.A., and Woods,V.L., Jr. (2004). Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS. Proc. Natl. Acad. Sci. U. S. A *101*, 751-756.

Parry,D.A. (1982). Coiled-coils in alpha-helix-containing proteins: analysis of the residue types within the heptad repeat and the use of these data in the prediction of coiled-coils in other proteins. Biosci. Rep. *2*, 1017-1024.

Paytubi,S., Madrid,C., Forns,N., Nieto,J.M., Balsalobre,C., Uhlin,B.E., and Juarez,A. (2004). YdgT, the Hha paralogue in *Escherichia coli*, forms heteromeric complexes with H-NS and StpA. Mol. Microbiol. *54*, 251-263.

Peat,T.S., Christopher,J.A., and Newman,J. (2005). Tapping the Protein Data Bank for crystallization information. Acta Crystallogr. D. Biol. Crystallogr. *61*, 1662-1669.

Petock,J.M., Wang,Y.F., DuBois,G.C., Harrison,R.W., and Weber,I.T. (2001). Effects of different post-crystallization soaking conditions on the diffraction of Mtcp1 crystals. Acta Crystallogr. D. Biol. Crystallogr. *57*, 763-765.

Pinson,V., Takahashi,M., and Rouviere-Yaniv,J. (1999). Differential binding of the *Escherichia coli* HU, homodimeric forms and heterodimeric form to linear, gapped and cruciform DNA. J Mol. Biol. *287*, 485-497.

Pollastri,G. and McLysaght,A. (2005). Porter: a new, accurate server for protein secondary structure prediction. Bioinformatics. *21*, 1719-1720.

Pollastri,G., Przybylski,D., Rost,B., and Baldi,P. (2002). Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. Proteins *47*, 228-235.

Prosseda,G., Falconi,M., Giangrossi,M., Gualerzi,C.O., Micheli,G., and Colonna,B. (2004). The *virF* promoter in Shigella: more than just a curved DNA stretch. Mol. Microbiol. *51*, 523-537.

Puglisi,J.D. and Tinoco,I., Jr. (1989). Absorbance melting curves of RNA. Methods Enzymol. *180*, 304-325.

Quiocho,F.A. and Richards,F.M. (1964). Intermolecular Cross Linking of a Protein in the Crystalline State: Carboxypeptidase-A. Proc. Natl. Acad. Sci. U. S. A *52*, 833-839.

Radaev,S., Li,S., and Sun,P.D. (2006). A survey of protein-protein complex crystallizations. Acta Crystallogr. D. Biol. Crystallogr. *62*, 605-612.

Renzoni,D., Esposito,D., Pfuhl,M., Hinton,J.C., Higgins,C.F., Driscoll,P.C., and Ladbury,J.E. (2001). Structural characterization of the N-terminal oligomerization domain of the bacterial chromatin-structuring protein, H-NS. J. Mol. Biol. *306*, 1127-1137.

Richey,B., Cayley,D.S., Mossing,M.C., Kolka,C., Anderson,C.F., Farrar,T.C., and Record,M.T., Jr. (1987). Variability of the intracellular ionic environment of *Escherichia coli*. Differences between *in vitro* and *in vivo* effects of ion concentrations on protein-DNA interactions and gene expression. J Biol. Chem. *262*, 7157-7164.

Rimsky,S., Zuber,F., Buckle,M., and Buc,H. (2001). A molecular mechanism for the repression of transcription by the H-NS protein. Mol. Microbiol. *42*, 1311-1323.

Robison,K., McGuire,A.M., and Church,G.M. (1998). A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete *Escherichia coli* K-12 genome. J Mol. Biol. *284*, 241-254.

Rost,B. (1996). PHD: predicting one-dimensional protein structure by profile-based neural networks. Methods Enzymol. *266*, 525-539.

Sali,A. and Blundell,T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J Mol. Biol. *234*, 779-815.

Sali,A. and Overington,J.P. (1994). Derivation of rules for comparative protein modeling from a database of protein structure alignments. Protein Sci. *3*, 1582-1596.

Sali,A., Potterton,L., Yuan,F., van,V.H., and Karplus,M. (1995). Evaluation of comparative protein modeling by MODELLER. Proteins *23*, 318-326.

Schneider,R., Travers,A., Kutateladze,T., and Muskhelishvili,G. (1999). A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*. Mol. Microbiol. *34*, 953-964.

Schroder,O., Tippner,D., and Wagner,R. (2001). Toward the three-dimensional structure of the *Escherichia coli* DNA-binding protein H-NS: A CD and fluorescence study. Biochem. Biophys. Res. Commun. *282*, 219-227.

Schuck,P. (1998). Sedimentation analysis of noninteracting and self-associating solutes using numerical solutions to the Lamm equation. Biophys. J *75*, 1503-1512.

Schuck,P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys. J *78*, 1606-1619.

Schuck,P. (2003). On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. Anal. Biochem. *320*, 104-124.

Shi,X. and Bennett,G.N. (1994). Plasmids bearing *hfq* and the *hns*-like gene *stpA* complement *hns* mutants in modulating arginine decarboxylase gene expression in *Escherichia coli.* J. Bacteriol. *176*, 6769-6775.

Shin,M., Song,M., Rhee,J.H., Hong,Y., Kim,Y.J., Seok,Y.J., Ha,K.S., Jung,S.H., and Choy,H.E. (2005). DNA looping-mediated repression by histone-like protein H-NS: specific requirement of Esigma70 as a cofactor for looping. Genes Dev. *19*, 2388-2398.

Shindo,H., Iwaki,T., Ieda,R., Kurumizaka,H., Ueguchi,C., Mizuno,T., Morikawa,S., Nakamura,H., and Kuboniwa,H. (1995). Solution structure of the DNA binding domain of a nucleoid-associated protein, H-NS, from *Escherichia coli.* FEBS Lett. *360*, 125-131.

Shindo,H., Ohnuki,A., Ginba,H., Katoh,E., Ueguchi,C., Mizuno,T., and Yamazaki,T. (1999). Identification of the DNA binding surface of H-NS protein from *Escherichia coli* by heteronuclear NMR spectroscopy. FEBS Lett. *455*, 63-69.

Sica,F., Demasi,D., Mazzarella,D.L., Zagari,A., Capasso,S., Pearl,L.H., D'Auria,S., Raia,C.A., and Rossi,M. (1994). Elimination of twinning in crystals of *Sulfolobus sofataricus* alcohol dehydrogenase holo-enzyme by growth in agarose gels. Acta Crystallogr. D Biol. Crystallogr. *50*, 508-511.

Slep,K.C., Rogers,S.L., Elliott,S.L., Ohkura,H., Kolodziej,P.A., and Vale,R.D. (2005). Structural determinants for EB1-mediated recruitment of APC and spectraplakins to the microtubule plus end. J Cell Biol. *168*, 587-598.

Smyth,C.P., Lundback,T., Renzoni,D., Siligardi,G., Beavil,R., Layton,M., Sidebotham,J.M., Hinton,J.C., Driscoll,P.C., Higgins,C.F., and Ladbury,J.E. (2000). Oligomerization of the chromatin-structuring protein H-NS. Mol. Microbiol. *36*, 962-972.

Sobott,F., McCammon,M.G., Hernandez,H., and Robinson,C.V. (2005). The flight of macromolecular complexes in a mass spectrometer. Philos. Transact. A Math. Phys. Eng Sci. *363*, 379-389.

Sonden,B. and Uhlin,B.E. (1996). Coordinated and differential expression of histone-like proteins in *Escherichia coli*: regulation and function of the H-NS analog StpA. EMBO J. *15*, 4970-4980.

Sonnenfield,J.M., Burns,C.M., Higgins,C.F., and Hinton,J.C. (2001). The nucleoid-associated protein StpA binds curved DNA, has a greater DNA-binding affinity than H-NS and is present in significant levels in *hns* mutants. Biochimie *83*, 243-249.

Sousa,R. (1995). Use of glycerol, polyols and other protein structure stabilizing agents in protein crystallization. Acta Crystallographica Section D *51*, 271-277.

Soutourina,O.A., Krin,E., Laurent-Winter,C., Hommais,F., Danchin,A., and Bertin,P.N. (2002). Regulation of bacterial motility in response to low pH in *Escherichia coli*: the role of H-NS protein. Microbiology *148*, 1543-1551.

Sperandio,V., Mellies,J.L., Delahay,R.M., Frankel,G., Crawford,J.A., Nguyen,W., and Kaper,J.B. (2000). Activation of enteropathogenic *Escherichia coli* (EPEC) LEE2 and LEE3 operons by Ler. Mol. Microbiol. *38*, 781-793.

Spurio,R., Durrenberger,M., Falconi,M., La,T.A., Pon,C.L., and Gualerzi,C.O. (1992). Lethal overproduction of the *Escherichia coli* nucleoid protein H-NS: ultramicroscopic and molecular autopsy. Mol. Gen. Genet. *231*, 201-211.

Spurio,R., Falconi,M., Brandi,A., Pon,C.L., and Gualerzi,C.O. (1997). The oligomeric structure of nucleoid protein H-NS is necessary for recognition of intrinsically curved DNA and for DNA bending. EMBO J. *16*, 1795-1805.

Steck,T.R., Pruss,G.J., Manes,S.H., Burg,L., and Drlica,K. (1984). DNA supercoiling in gyrase mutants. J Bacteriol. *158*, 397-403.

Stella,S., Spurio,R., Falconi,M., Pon,C.L., and Gualerzi,C.O. (2005). Nature and mechanism of the *in vivo* oligomerization of nucleoid protein H-NS. EMBO J. *24*, 2896-2905.

Suzuki,T., Ueguchi,C., and Mizuno,T. (1996). H-NS regulates OmpF expression through *micF* antisense RNA in *Escherichia coli*. J. Bacteriol. *178*, 3650-3653.

Svedberg,T. and Pedersen,K.O. (1940). The Ultracentrifuge. Oxford University Press, London).

Swinger,K.K., Lemberg,K.M., Zhang,Y., and Rice,P.A. (2003). Flexible DNA bending in HU-DNA cocrystal structures. EMBO J *22*, 3749-3760.

Tendeng,C. and Bertin,P.N. (2003). H-NS in Gram-negative bacteria: a family of multifaceted proteins. Trends Microbiol. *11*, 511-518.

Tendeng,C., Krin,E., Soutourina,O.A., Marin,A., Danchin,A., and Bertin,P.N. (2003). A Novel H-NS-like protein from an antarctic *psychrophilic* bacterium reveals a crucial role for the N-terminal domain in thermal stability. J. Biol. Chem. *278*, 18754-18760.

Tippner,D., Afflerbach,H., Bradaczek,C., and Wagner,R. (1994). Evidence for a regulatory function of the histone-like *Escherichia coli* protein H-NS in ribosomal RNA synthesis. Mol. Microbiol. *11*, 589-604.

Tippner,D. and Wagner,R. (1995). Fluorescence analysis of the *Escherichia coli* transcription regulator H-NS reveals two distinguishable complexes dependent on binding to specific or nonspecific DNA sites. J. Biol. Chem. *270*, 22243-22247.

Trakhanov,S. and Quiocho,F.A. (1995). Influence of divalent cations in protein crystallization. Protein Sci. *4*, 1914-1919.

Travers,A. and Muskhelishvili,G. (2005). Bacterial chromatin. Curr. Opin. Genet. Dev. *15*, 507-514.

Tupper,A.E., Owen-Hughes,T.A., Ussery,D.W., Santos,D.S., Ferguson,D.J., Sidebotham,J.M., Hinton,J.C., and Higgins,C.F. (1994). The chromatin-associated protein H-NS alters DNA topology *in vitro*. EMBO J. *13*, 258-268.

Ueguchi,C., Kakeda,M., and Mizuno,T. (1993). Autoregulatory expression of the *Escherichia coli hns* gene encoding a nucleoid protein: H-NS functions as a repressor of its own transcription. Mol. Gen. Genet. *236*, 171-178.

Ueguchi,C. and Mizuno,T. (1993). The *Escherichia coli* nucleoid protein H-NS functions directly as a transcriptional repressor. EMBO J. *12*, 1039-1046.

Ueguchi,C., Seto,C., Suzuki,T., and Mizuno,T. (1997). Clarification of the dimerization domain and its functional significance for the *Escherichia coli* nucleoid protein H-NS. J. Mol. Biol. *274*, 145-151.

Ueguchi,C., Suzuki,T., Yoshida,T., Tanaka,K., and Mizuno,T. (1996). Systematic mutational analysis revealing the functional domain organization of *Escherichia coli* nucleoid protein H-NS. J. Mol. Biol. *263*, 149-162.

Umanski,T., Rosenshine,I., and Friedberg,D. (2002). Thermoregulated expression of virulence genes in enteropathogenic *Escherichia coli*. Microbiology *148*, 2735-2744.

Van Den Heuvel,R.H. and Heck,A.J. (2004). Native protein mass spectrometry: from intact oligomers to functional machineries. Curr. Opin. Chem. Biol. *8*, 519-526.

van Noort J., Verbrugge,S., Goosen,N., Dekker,C., and Dame,R.T. (2004). Dual architectural roles of HU: formation of flexible hinges and rigid filaments. Proc. Natl. Acad. Sci. U. S. A *101*, 6969-6974.

Varshavsky,A.J., Nedospasov,S.A., Bakayev,V.V., Bakayeva,T.G., and Georgiev,G.P. (1977). Histone-like proteins in the purified *Escherichia coli* deoxyribonucleoprotein. Nucleic Acids Res. *4*, 2725-2745.

Vidal,O., Robert,M.C., Arnoux,B., and Capelle,B. (1999). Crystalline quality of lysozyme crystals grown in agarose and silica gels studied by X-ray diffraction techniques. Journal of Crystal Growth *196*, 559-571.

Waldsich,C., Grossberger,R., and Schroeder,R. (2002). RNA chaperone StpA loosens interactions of the tertiary structure in the td group I intron *in vivo*. Genes Dev. *16*, 2300-2312.

Wang,S., Cosstick,R., Gardner,J.F., and Gumport,R.I. (1995). The specific binding of *Escherichia coli* integration host factor involves both major and minor grooves of DNA. Biochemistry *34*, 13082-13090.

Watenpaugh,K.D. (1985). Overview of phasing by isomorphous replacement. Methods Enzymol. *115*, 3-15.

Waterman,S.R. and Small,P.L. (2003). Transcriptional expression of *Escherichia coli* glutamate-dependent acid resistance genes *gadA* and *gadBC* in an *hns rpoS* mutant. J. Bacteriol. *185*, 4644-4647.

Wattenberg,A., Sobott,F., and Brutschy,B. (2000). Detection of intact hemoglobin from aqueous solution with laser desorption mass spectrometry. Rapid Commun. Mass Spectrom. *14*, 859-861.

Williams,R.M. and Rimsky,S. (1997). Molecular aspects of the *E. coli* nucleoid protein, H-NS: a central controller of gene regulatory networks. FEMS Microbiol. Lett. *156*, 175-185.

Williams,R.M., Rimsky,S., and Buc,H. (1996). Probing the structure, function, and interactions of the *Escherichia coli* H-NS and StpA proteins by using dominant negative derivatives. J. Bacteriol. *178*, 4335-4343.

Williamson,H.S. and Free,A. (2005). A truncated H-NS-like protein from enteropathogenic *Escherichia coli* acts as an H-NS antagonist. Mol. Microbiol. *55*, 808-827.

Winzor,D.J. (2003). Analytical exclusion chromatography. J. Biochem. Biophys. Methods *56*, 15-52.

Woldringh,C.L., Jensen,P.R., and Westerhoff,H.V. (1995). Structure and partitioning of bacterial DNA: determined by a balance of compaction and expansion forces? FEMS Microbiol. Lett. *131*, 235-242.

Xu,J. and Johnson,R.C. (1995). Identification of genes negatively regulated by Fis: Fis and RpoS comodulate growth-phase-dependent gene expression in *Escherichia coli*. J Bacteriol. *177*, 938-947.

Yadav,M.K., Leman,L.J., Price,D.J., Brooks,C.L., III, Stout,C.D., and Ghadiri,M.R. (2006). Coiled coils at the edge of configurational heterogeneity. Structural analyses of parallel and antiparallel homotetrameric coiled coils reveal configurational sensitivity to a single solvent-exposed amino acid substitution. Biochemistry *45*, 4463-4473.

Yang,S.W. and Nash,H.A. (1995). Comparison of protein binding to DNA *in vivo* and *in vitro*: defining an effective intracellular target. EMBO J *14*, 6292-6300.

Yoshimura,S.H., Hizume,K., Ohniwa,R.L., Ishihama,A., and Takeyasu,K. (2004). Fundamental structural units of the *Escherichia coli* nucleoid revealed by atomic force microscopy. Nucleic Acid Research *32*, 1982-1992.

Zhang,A. and Belfort,M. (1992). Nucleotide sequence of a newly-identified *Escherichia coli* gene, *stpA*, encoding an H-NS-like protein. Nucleic Acids Res. *20*, 6735.

Zhang,A., Rimsky,S., Reaban,M.E., Buc,H., and Belfort,M. (1996). *Escherichia coli* protein analogs StpA and H-NS: regulatory loops, similar and disparate effects on nucleic acid dynamics. EMBO J. *15*, 1340-1349.

Zhao,X., Ghaffari,S., Lodish,H., Malashkevich,V.N., and Kim,P.S. (2002). Structure of the Bcr-Abl oncoprotein oligomerization domain. Nat. Struct. Biol. *9*, 117-120.

Zimmerman,S.B. (2006). Shape and compaction of *Escherichia coli* nucleoids. J Struct. Biol. *156*, 255-261.

Zuber,F., Kotlarz,D., Rimsky,S., and Buc,H. (1994). Modulated expression of promoters containing upstream curved DNA sequences by the *Escherichia coli* nucleoid protein H-NS. Mol. Microbiol. *12*, 231-240.